WORLD VIEW Women scientists should embrace media opportunities **p.733**

GLOW Scorpions avoid dancing in the moonlight **p.734**

SUGAR Legal block on beet route around ban on GM plants **p.736**

Singapore's salad days are over

Uncertainty has replaced confidence as economic reality bites science in the city-state and scientists find that their research funds now come with strings attached.

When Neal Copeland and Nancy Jenkins, a renowned husband-and-wife team of cancer geneticists, left the US National Cancer Institute in Bethesda, Maryland, for the Institute of Molecular and Cell Biology in Singapore in 2006, they joined a string of star names in the city-state that suggested its remarkable investment in research was paying off. Generous funds have flowed to science in Singapore for the best part of a decade, and researchers from around the world have followed. Drawn by hefty salaries and enviable working conditions, they have rapidly given Singapore an international presence. The Genome Institute of Singapore, for example, has asserted itself as one of the most important basic genomics research organizations in the world.

Best of all for scientists, despite Singapore's reputation for top-down autocracy, its investment in research came with relatively few strings attached. The administration realized that researchers prefer to have the freedom to follow their curiosity and that, to attract the best minds, they needed to loosen the reins. As a result, Singapore's biomedical infrastructure seems set to enter the next stage in its development, in which researchers looking for their next posts — especially the much-sought promising young researchers and postdoctoral students — are starting to consider Singapore, not only because of the large grants, but also because of its scientific reputation and intellectual ferment.

To many outsiders, the Singapore experiment seemed too good to be true — and perhaps it was. Singapore is not immune to the economic pressure mounting on research communities around the world, and policy-makers everywhere want returns on their investments. Rumours of purse-tightening measures have grown over the past year, but researchers in the city-state were still stunned by the news in September that almost one-third of the total research budget will be abruptly shifted to competitive 'industrial alignment funds'. Access to that funding will now depend on researchers' abilities to show that their work has industrial applications. The policy will affect all research but is aimed particularly at the biomedical sciences, which are senior figures feel are not pulling their weight.

Nobody should cry for Singapore's scientists, who don't expect sympathy. They have been living large and will continue, if they can prove themselves, to be paid generously. And having to write grant applications is not enslavement — it is the norm for most researchers around the world. The problem is not Singapore's shifting priorities, but how the government is implementing the change.

In response to a call for research proposals last month, Singapore's scientists have had to scramble to draft application-oriented proposals. They know that industrial contracts would help. But, given the shaky state of the global pharmaceutical industry, such contracts are not easy to come by. Many applications are going in with a weak note: "industrial partner to be decided". Singapore's scientists worry that, given only weeks or months to secure deals, they will be forced into unfavourable agreements. One researcher at Singapore's Agency for Science, Technology and Research says that the policy is an attempt to turn the agency "into a contract-research organization overnight".

Researchers also worry that the government has not made clear how it will review the sudden influx of research applications. Singapore has used external review committees to audit its institutes in the past. But reviewing individual grants is a different and much more labour-intensive procedure if done properly. Will Singapore be forced to rely on a small number of bureaucrats and selected scientists for reviews? Frustrated by the changes, Copeland and Jenkins have decided to leave Singapore. Many other scientists there are also looking for new posts.

"To many outsiders, the Singapore experiment seemed too good to be true — and perhaps it was."

The government should move quickly to clarify the grant-review process. Easing the industrial-application restrictions would help scientists in the short term. More fundamentally, as researchers have suggested, the government could phase in the funding changes over the next few years, rather than introducing them all at once.

Singapore's rapid transformation came about through massive, perhaps even excessive, funding. The move to align scientific objectives with economic reality is understandable. But it would be a huge waste if doing so with undue haste and insufficient planning were to destroy Singapore's impressive experiment. ■

Animal instinct

Germany must better explain the scientific use of animals to remain a major biomedical force.

Ten years ago, researchers using animals in Britain found themselves in a particularly hostile environment. A campaign of intimidation and violence by animal-rights extremists had spun out of control. The London-based lobby group Understanding Animal Research — a historic organization founded in 1908 — responded with a counter-campaign of its own that, in 2005, smoothed the introduction of laws giving the police increased powers to stop extremists from harassing scientists and from harming animal-research organizations.

Scientists in Germany have not yet experienced such a degree of violence, although the potential is there. In one incident in Munich, activists rented billboard space to display the name, home address and telephone number of a scientist whose research involved animals. In another, they distributed flyers describing a local researcher as a killer and torturer. Similar or worse incidents have occurred in other cities such as Bremen

and Tübingen, where biomedical research is particularly strong.

Biomedical scientists in Germany perceive a separate crisis — increasing legislative restrictions that make it more difficult to carry out animal experiments. Hearing little to the contrary from researchers themselves, the public tends to assume that animal experiments are an unnecessary evil, so politicians respond with more restrictions.

That problem was a major motivation for the Basel Declaration — drafted and signed at a meeting in Basel, Switzerland, last week (see page 742). Its signatories pledge to engage in open debate with the public about their work on animal experiments, to stress the high ethical standards to which they adhere and to explain why they have to do it. They intend, for example, to visit local schools or to mention that their research used animals when speaking to the press about new results. Germany does not have a national organization such as Understanding Animal Research to manage and maintain this public outreach. Now is the time that it should.

During the UK crisis, Understanding Animal Research used the momentum of the campaign against extremist violence to engage British scientists to establish a public dialogue. Education on the medical value of animal research helped to dismantle knee-jerk public sympathy for animal-rights campaigns and encouraged politicians to act.

In the 1990s, the pharmaceutical industry largely abandoned Germany as a research base, put off by restrictions on genetic technologies and the use of animals. Today the country is a leader in biomedical research, and red tape around genetic technologies has been significantly reduced. Yet the animal issue remains sensitive. Scientists there have tended to keep their heads down and hope for the best. But they should fear the worst: a crisis such as that in Britain could arise at any time. Some of the five main German research organizations, such as the Max Planck Society, which runs 80 research institutes in different disciplines, and the DFG, Germany's main research-granting agency, acknowledge the animal-experiment problem and have small offices that monitor legislative activities. But they do not engage in significant public outreach.

The solution must be a single, non-partisan national office that can implement the principles of the Basel Declaration. It need not be large — Understanding Animal Research has only nine staff — but it needs to be professional. Busy researchers do not have the time or the lobbying skills to organize long-term concerted action. Who should pay? When it comes to the defence of research in Germany in general, the research organizations and universities band together as the formidable Alliance of German Science Organizations — the 'Allianz'. Successive governments have deferred to it, and have committed to long-term funding increases even in times of financial crisis.

The animal issue has an intrinsically emotive nature and is never going to go away.

The Allianz is the appropriate body to create and fund a German organization analogous to Understanding Animal Research. Including industry might breed distrust. The Basel Declaration has shown that animal researchers in Germany are willing to go public. A small investment by all members of the Allianz would bolster this new solidarity and serve as insurance for Germany's biomedical effort and for the new biomedical industries that are springing up.

The animal issue has an intrinsically emotive nature and is never going to go away. To keep the public ignorant of the benefits of animal research — without which it is currently impossible to develop any new therapies — was never a solution. ■

Give up the ghosts

Funding agencies should make researchers reveal industry links.

The spectral fingerprints of a big drug company have once again been found all over academic publications. Documents released last week by a watchdog group based in Washington DC raise concerns about the role of writers paid by GlaxoSmithKline (GSK) in works attributed to psychiatric researchers at a number of US institutions. They add to the drumbeat of allegations in recent years indicating that such ghostwriting — in which articles contain substantial portions written by someone who is not listed as an author — is endemic in the biomedical literature.

The documents were made available as a result of litigation over GSK's antidepressant Paxil (paroxetine) and were pounced on by the Project on Government Oversight, which raised concerns about authorship of a research article, journal editorial and textbook.

The researchers did acknowledge the alleged ghostwriters of the textbook and the editorial in notes, but only for "editorial support". For the journal article, which appeared in a supplement to *Psychopharmacology Bulletin*, GSK is thanked for an "unrestricted educational grant". But the original front page of the manuscript — which the academic author is instructed to remove before submission to the journal — declares that it was prepared by writers from Scientific Therapeutics Information, a company based in Springfield, New Jersey, hired by GSK. The article and textbook discuss the uses of Paxil. The editorial, in *Biological Psychiatry*, gives an overview of depression as a major and growing public-health problem — which certainly does no harm to a company aggressively marketing an antidepressant.

The academic authors and the American Psychiatric Association, which published the textbook, have strongly denied that the pharmaceutical giant influenced its content. So, too, have the authors of the editorial and the journal article. GSK shareholders, then, may wonder what the company got for its money. The issue here is not that industry-financed experts cannot write useful and unbiased reports, but that their role must be declared in full. It is for readers, not authors, to conclude that there is no conflict of interest.

All the academic authors involved in this case have been recipients of US National Institutes of Health (NIH) funding; all but one still are. The NIH may argue, rightly, that the ghostwritten publications did not use its money. It will also note, correctly, that this in an issue that demands far broader action. Both are beside the point. Money is fungible, and rarely do the studies and intellectual output of senior researchers divide neatly into industry-funded and taxpayer-funded work. If its grantees are not playing by the rules, the NIH is tarred and public trust is damaged. So, how clear are the rules on ghostwriting? A study last year found that just 10 out of 50 top US academic medical centres had explicit, web-accessible policies that prohibit the practice. Another three banned ghostwriting in practice without naming it as such (J. R. Lacasse and J. Leo *PLoS Med.* **7**, e1000230; 2010).

Discussing the issue of ghostwriting a year ago, Francis Collins, the NIH director, said publicly that he was "shocked" that "people would allow their names to be used on articles they did not write, that were written for them, particularly by companies that have something to gain by the way the data is presented". Many will share that shock, but, unlike Collins, few are in a position to do something about it. The agency is "considering how best to address and ensure" greater transparency and accountability as its grantees develop and author articles, Sally Rockey, NIH chief of extramural research, told *Nature* in an e-mail last week.

A good start would be for the NIH to require all institutions that take its funds to articulate, publicize and vigorously enforce a clear ban on ghostwriting. Other funders should follow suit. Without such a clear signal, and the willingness to give a ban teeth, this troubling ghost will linger at the feast. ■

➲ NATURE.COM
To comment online, click on Editorials at:
go.nature.com/xhunqv

WORLD VIEW *A personal take on events*

R. P. GRANT

Women scientists must speak out

Female researchers still battle sexism. The media gives them an opportunity to be heard alongside male colleagues, says **Jennifer Rohn**.

Despite decades of awareness, science is still inherently sexist. Women are vastly under-represented in professorships and in national academies worldwide. This is a familiar problem, but less highlighted is how the discrepancy plays out in the public arena of science — the media.

Male science pundits dominate television, radio and print — including the pages of opinion and comment in this journal. This imbalance cannot simply be explained by the shortage of female professors, as many male pundits are still at an early stage of their academic careers, when genders are better balanced. So what is behind this effective invisibility of women scientists in our media? And why does it matter?

Many people think that women themselves are to blame for the male-dominated media, in science and other subjects. Women, who often bear the brunt of domestic obligations, are said to have less time than men to participate in activities outside their work. And female colleagues tend to say that they do not feel eminent or qualified enough to comment. Perhaps this points to a question of confidence — one that does not seem to bother most men. Women may also be uncomfortable with the cut and thrust of conflict and debate. Indeed, at scientific seminars I have attended, most of the questions come from men, despite the audience usually containing an equal number of women. Voicing one's opinion in a public arena is a charged activity that seems to discourage many women, yet this is precisely the skill that a good pundit needs.

POWERS OF ANALYSIS AND PERSUASION ARE USEFUL SKILLS FOR ALL SCIENTISTS, REGARDLESS OF SEX.

This still cannot explain the near-total absence of women pundits. Sexism must be responsible too. Having both the inclination and the time to do media work myself, I have certainly found myself dropped for programmes and replaced by less-qualified men. A prominent television producer once refused to put a colleague on screen because, he said, people wouldn't swallow science offered from "a young, blonde girl" like her. I have voiced opinions during panel discussions to little effect, then watched a man next to me say the same thing to widespread applause. In group discussions, I find that women are often talked over by men as if they weren't even in the room, whereas men are more likely to let other men finish their sentences. More insidiously, it is well documented that what passes for spirited assertion in men is interpreted, by both sexes, as unpleasant aggression in women. Given this bias, I understand why many women might prefer not to get involved.

Although these external factors and biases are out of our control, there are positive steps that women scientists can take to increase their visibility in the public arena. First and most importantly, women need to speak up. They could start in the relative safety of their own academic departments, preferably during their PhD studies. It is not easy — a famous female professor recently admitted to me that she still gets palpitations when asking questions at high-profile academic seminars — but nerves never killed anyone. Work through them, and you will gain respect as someone who has intelligent things to say and is not afraid to share them. Verbal sparring at seminars can also help your career because it builds confidence, develops an ability to communicate ideas and can even lead to collaborations. (And, palpitations aside, it gets easier with practise.) From speaking out at seminars, I found it natural to progress to media work, which, as well as being challenging and enjoyable, hones your powers of analysis and persuasion — skills that are useful for all scientists, regardless of sex.

Second, keep in mind that, to the media and its audience, you don't have to be an eminent professor to have a valuable opinion — any PhD student or postdoc is miles ahead of the public in terms of scientific knowledge. Start a blog about your own research to refine your opinions and develop a style. As you gain more research experience, give your name and telephone number to your institution's press office, and don't shy away if asked. Similarly, don't be afraid to stray from your specialized niche of research expertise: if you are reasonably well read on a general topic, your opinion will still be useful. It is important to participate, because if we scientists aren't ready to step into the gap at short notice, the press may choose someone who isn't qualified at all — a real problem when the story is about homeopathy or other quackery.

Some might question if it matters whether we have more female science pundits, as long as the men are doing the job well. I think it does. A female messenger could attract a more diverse crowd, including other women. The point of punditry is often to persuade people that science is worthwhile and, more to the point, deserves funding. Also, pundits help to put forward scientific recommendations and counter misinformation. When it comes to controversial issues such as climate change, childhood vaccinations or genetically modified food, we need as many people as possible to hear and engage with our arguments. Women should stand shoulder to shoulder with their male colleagues to make this happen. ■

➲ NATURE.COM
Discuss this article online at:
go.nature.com/hthvsy

Jennifer Rohn *is a cell biologist at University College London and editor of LabLit.com. Her most recent book is* The Honest Look *(Cold Spring Harbor Laboratory Press).*
e-mail: jenny@lablit.com

RESEARCH HIGHLIGHTS *Selections from the scientific literature*

ASTRONOMY

Stars reach their limits

Stars seem to have a size limit of 100–150 times that of the Sun; the reason for this has been the subject of debate. Some astronomers think that a large star's own radiation blows away the gas it needs to grow, whereas others suggest that a star's progenitor cloud fragments.

A new simulation bolsters the fragmentation theory. Thomas Peters at the University of Heidelberg in Germany and his colleagues simulated the birth of massive stars in a cloud of gas and found that other, smaller stars formed from the fragmenting gas before the largest one could grow too big. The simulations are supported by some observations, and should lead to a better understanding of how big stars form.

Astrophys. J. 725, **134–145 (2010)**

H. BORTMAN

MICROBIOLOGY

Bacteria that thrive on arsenic

A bacterium discovered in a lake high in arsenic not only metabolizes the normally toxic element, but also seems to incorporate it into its DNA and other molecules in place of phosphorus. This hints at a biochemistry very different from that long thought to underlie life on Earth.

Felisa Wolfe-Simon at the US Geological Survey in Menlo Park, California, and her colleagues found the microbe in California's Mono Lake (pictured). When cultured in arsenate with only trace amounts of phosphate, the organism grew at a rate equal to 60% of that it achieves in phosphate.

Using radiolabelling and mass spectroscopy, the team found arsenic in cellular fractions of the bacterium's proteins, lipids, metabolites and nucleic acids in amounts similar to those expected for phosphate in normal cell biochemistry. X-ray analysis suggested that the arsenic takes the form of arsenate, and bonds with carbon and oxygen similarly to phosphate.

Science **doi:10.1126/science.1197258 (2010)**

For details of the mixed reactions to this surprising finding see p.741.

ZOOLOGY

Scorpions glow to sense

Scorpions' fluorescence under ultraviolet (UV) light may help them to detect and avoid the light. Because night-time levels of UV light correlate with the Moon phase, this could enable the creatures to detect moonlight and remain obscured on moonlit nights.

Carl Kloock and his team at California State University in Bakersfield reduced the glow of 15 female *Paruroctonus becki* scorpions (**pictured**) by exposing them to 16 hours of low-level UV light per day. The authors placed the creatures, along with 15 control, fluorescing scorpions, in Petri dishes that were painted black across one half. The scorpions were then exposed to infrared (IR) light only, IR and UV, or IR and white light.

The team found that, when exposed to UV light, the fluorescent scorpions were less active than the reduced-fluorescent ones, moving less often between the light and dark parts of the Petri dishes.

J. Arachnol. 38, **441–445 (2010)**

J. MARC-MOHAMED

MATERIALS SCIENCE

Antiseptic silver slivers

Silver is toxic to bacteria, and nanoparticles of the element offer promise as a coating for medical devices. But silver nanoparticles readily oxidize and clump together, losing their antibacterial activity. Chunhai Fan at the Chinese Academy of Sciences in Shanghai and his colleagues solved this problem by growing the particles on biocompatible silicon nanowires. This avoids the need for toxic or expensive chemicals to stabilize the silver nanoparticles.

The researchers show that exposure to a 10% solution of silver-coated nanowires froze population size in the bacteria *Escherichia coli* and *Bacillus subtilis* throughout a two-day test period.

Adv. Mater. **doi:10.1002/adma.201001934 (2010)**

OPTICS

Taking wing on a beam of light

A beam of light has been used to provide lift to a micrometre-sized curved rod in a manner analogous to that by which air passing over a wing provides lift to birds and aeroplanes.

Grover Swartzlander at the Rochester Institute of Technology in New York and his colleagues shone a weakly focused laser beam through the roughly semi-cylindrical rods, which refracted the light rays. This refraction changed the direction of the rays' momentum, causing an equal and opposite momentum change on the rods themselves. Because of the rods' asymmetrical shape, the momentum shift was directed more towards one side, driving the rods upwards at around 2.5 micrometres per second.

The researchers suggest that the technique could be used to transport microscopic machines through liquids and to help steer solar sails in spacecraft.

***Nature Photon.* doi:10.1038/nphoton.2010.266 (2010)**

For a longer story on this research, see go.nature.com/ye4zid.

NEUROSCIENCE

Enzyme helps pain persist

Pain perception in mice is maintained for several days by augmented activity of a particular enzyme in a brain area associated with chronic pain. Blocking the enzyme, PKMζ, with a peptide inhibitor alleviates the pain.

Bong-Kiun Kaang at Seoul National University, Min Zhuo at the University of Toronto in Canada and their colleagues show that in the days following a nerve injury, mice had higher levels of PKMζ in a brain region known as the anterior cingulate cortex (ACC). Injecting the PKMζ inhibitor into the ACC led to a drop in synaptic activity, or neuronal communication, in that region, as well as a decrease in pain responses.

The authors suggest that PKMζ mediates chronic pain by boosting synaptic transmission in the ACC.

***Science* 330, 1400–1404 (2010)**

ECOLOGY

Reptiles rose after forests died

The disappearance of vast tracts of tropical forest some 305 million years ago led to an explosion in the global diversity of reptiles and amphibians, thanks to the emergence of many new, fragmented habitats.

Howard Falcon-Lang at Royal Holloway, University of London, in Surrey, UK, and his colleagues compared the distribution and diversity of these animals in the fossil record. During the period they studied, climate change dried up equatorial rainforests in the land mass that later became Europe and North America.

Many of the species that lived across these forests became extinct, and were replaced by a wealth of different types of reptile and amphibian that were particular to isolated habitats. Amphibians, which depend on aquatic environments, fared less well than reptiles, which were able to adapt to a drier world.

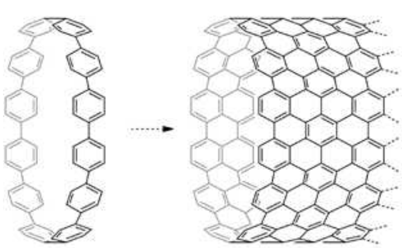
***Geology* 38, 1079–1082 (2010)**

CHEMISTRY

Ring-a-ring o' benzene

Loops of interconnected benzene rings have long fascinated chemists, who have now developed a more flexible way to string the rings together.

Kenichiro Itami and his co-workers at Nagoya University in Japan created the loop, or cycloparaphenylene (CPP, **pictured** below left), by first coupling L-shaped and linear molecules to form U-shaped ones. They then combined two of these molecules to form the desired O-shaped CPP.



WILEY

This modular approach has potential for producing specific CPPs of any size greater than 13 benzene rings. The loops could be useful for fabricating single-walled carbon nanotubes of specific diameter, by growing the loops horizontally (**pictured** right).

***Angew. Chem. Int. Edn* doi:10.1002/anie.201005734 (2010)**

COMMUNITY CHOICE

The most viewed papers in science

CANCER

Full immunity needed to fight cancer

HIGHLY READ on www.cell.com in November

Certain targeted cancer drugs shrink tumours by shutting down key genes. But researchers report that this may not be enough to vanquish cancer — a functional immune system is also a pre-requisite.

Immune cells are known to be important in restricting tumour formation, but less is known about their role in tumour regression. Dean Felsher at Stanford University in California and his team switched off genes required for tumour growth in mouse models of lymphoma and leukaemia. They found that the rate of tumour shrinkage fell when the mice lacked an intact immune system — to up to one-thousandth of the normal speed — and the frequency of tumour recurrence rose.

The team discovered that immune cells called $CD4^+$ T cells are needed to shut down blood-vessel growth and to trigger tumour-cell senescence. Moreover, a protein produced by the T cells called thrombospondin 1, which blocks blood-vessel formation, seems to be key to fending off tumours.

***Cancer Cell* 18, 485–498 (2010)**

EVOLUTION

Speaking in borrowed tongues

Languages evolve in a similar way to biological organisms, with ancestral languages splitting into descendent ones. In language evolution, 'lexical borrowing', whereby a word is transferred from one language to another, is also common. Linguists have struggled to distinguish between words that have descended and those that have been borrowed.

Tal Dagan at the Heinrich Heine University in Düsseldorf, Germany, and her colleagues looked for instances of borrowing by analysing the relationships between 2,346 words of basic vocabulary with similar meanings from 84 Indo–European languages. By studying networks of related words, the researchers found that, on average, 8% of the basic vocabulary in each of the languages is borrowed. Basic vocabulary was previously assumed to be fairly immune to borrowing.

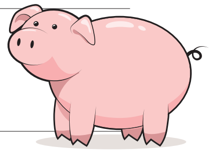
***Proc. R. Soc. B* doi: 10.1098/rspb.2010.1917 (2010)**

CORRECTION

In "The source of sour taste" (*Nature* **468,** 603; 2010), the mice with tagged bitter, sweet and umami taste cells were tested by Liman *et al.* but engineered by another lab. Furthermore, in response to acids, sour-taste cells did not conduct sodium ions, which were previously thought to mediate sour sensing.

NATURE.COM

For the latest research published by *Nature* visit:
www.nature.com/latestresearch



P. NICKLEN/NATL GEOGRAPHIC/GETTY

Nikumaroro Island, part of the Republic of Kiribati, is surrounded by waters rich in tropical tuna species.

FISHERIES

Islands champion tuna ban

Pacific nations to restrict fishing across a vast swathe of international waters.

BY CHRISTOPHER PALA

A bold move by eight Pacific island nations to preserve the world's last large stocks of tuna is expected to face strong resistance this week at a meeting of the Western and Central Pacific Fisheries Commission (WCPFC) near Honolulu, Hawaii.

By leveraging agreements with foreign fishermen in their own territorial waters, the islands have banned fleets that fish with purse-seine nets — mechanized nets that can capture entire schools of tuna in a single haul — from operating in a region of international waters roughly the size of India. The area, known as the Eastern High Seas (see map, overleaf), will still be fished by hook-and-line, which is considered biologically more sustainable. The islands will also cut the time that purse-seiners can spend fishing in their territorial waters by nearly a third. The restrictions, agreed to by the eight nations in April, are scheduled to take effect on 1 January 2011.

Marine biologists say the development is a major step forward for efforts to halt the global decline of bigeye, yellowfin, skipjack and other tropical tuna species. In October this year, Britain turned the entire Exclusive Economic Zone around the Chagos Islands, in the centre of the Indian Ocean, into a no-take zone, making it the first area rich in tuna that has been closed to fishing. At 3.2 million square kilometres, the Eastern High Seas is six times larger.

"These are the most far-reaching ocean-conservation measures ever," says Daniel Pauly, a leading fisheries scientist at the University of British Columbia in Vancouver. "For the first time since man has been fishing out the open oceans, we're going to see a reversal of the decline of pelagic species in two big areas."

➲ NATURE.COM
For more on fisheries and the challenges they face, see:
go.nature.com/mckets

At the meeting this week, the world's major fishing nations, including the United States, are expected to challenge the measures. By treaty, the United States is technically exempt from the restrictions, but two years ago it chose to side with the island nations in the closing of two smaller pockets of international waters to foreign fleets.

This time, conference sources predict a less sympathetic attitude. "We're not totally settled in our positions," says Charles Karnella of the National Oceanic and Atmospheric Administration, who heads the US delegation. "We're renegotiating our treaty and how the closure is dealt with will be part of our talks."

Ships from the United States operate under the South Pacific Tuna Treaty, in which the ▶

A. SEALE/IMAGEQUESTMARINE.COM

TUNA FACTS
Bigeye at risk

Second only to the dwindling bluefin in economic importance and vulnerability, the bigeye tuna (*Thunnus obesus*) is the deep diver of the tuna family. It typically spends most of its day foraging for fish and squid at depths of several hundred metres. At night, it rises closer to the surface, following its prey. Adults can reach 2.5 metres long and weigh up to 180 kilograms.

Prized for the texture, taste and colour of its meat, the bigeye is replacing bluefin as the most expensive tuna for sushi.

Adults are caught by hook-and-line vessels in numbers that researchers say are sustainable. But teenage bigeyes like to swim with schools of the same-sized, adult skipjack when these assemble in huge schools around fish-aggregating devices used by the purse-seine fleet. All end up in cans.

Although juvenile bigeyes are not preferred by canneries, purse-seiners use the devices to increase their take of skipjack.

Environmentalists and Pacific islanders have called for the devices to be banned to spare the bigeyes.

▶ US government pays most of their licence fees and provides US$18 million in foreign aid to 14 island nations. In exchange, the US fishing fleet, now set at 40 ships, has unlimited access to the region.

Should the United States decide not to abide by the new closures, US purse-seiners could find themselves fishing there virtually alone, contributing to the depletion of bigeye tuna at a time when fisheries scientists are calling for a 30% reduction in bigeye catch to avoid the collapse of stocks, which have fallen from 1.2 million tonnes to 500,000 tonnes since 1952 (see 'Bigeye at risk').

MARINE HAVENS

Sari Tolvanen, a Greenpeace International oceans campaigner attending the meeting, points out that the closure area is bordered on both sides by no-take areas created when former US president George W. Bush named the islands of Wake, Johnston, Jarvis, Howland, Baker and Palmyra Atoll as a Marine National Monument in January 2009.

"It would be shameful if the Obama administration did not follow the Bush administration's example and opted out of the conservation measures taken by the Pacific island nations," she says.

The decision to end purse-seining in the Eastern High Seas was taken by the Parties to the Nauru Agreement (PNA), eight Pacific island nations in whose waters 80% of the region's tuna is fished. They comprise the Federated States of Micronesia, Kiribati, the Marshall Islands, Nauru, Palau, Papua New Guinea, the Solomon Islands and Tuvalu.

Although no nation can legally restrict fishing in the high seas, the PNA countries have jointly amended the standard contracts they sign with foreign fleets with a stipulation that the fleets refrain from fishing in some international waters, in order to remain eligible for licences within waters directly controlled by the islands.

The regulations are made enforceable by the use of radio transponders, which reveal the positions of the licensed ships at all times. The first such measures, which began in 2008, closed smaller pockets of international waters that were being used as refuges for vessels fishing illegally. The extension of the ban to the Eastern High Seas is seen as potentially far more important because this region is large enough to have an impact on species preservation.

Officials from the PNA say that once they have succeeded in ending purse-seining in the Eastern High Seas, they will ban longliners — a different type of fishing vessel that catches about the same tonnage as a purse-seiner — from the area, effectively turning it into the world's largest marine reserve.

An important factor that will determine the effectiveness of the ban is the extent to which fish stay within the protected area. John Hampton, head of the fisheries programme at the Secretariat of the Pacific Community (SPC), co-authored a study conducted farther to the west. This found that half the skipjack tuna there spend their entire lives within a radius of 675–750 kilometres (J. Sibert & J. Hampton *Mar. Pol.* **27,** 87–95; 2003). A new SPC study under way in the Central Pacific focuses on bigeye tuna, with 15,000 tuna tagged so far, he says. It will provide information on tuna movement and mixing with adjacent areas, rates of mortality and other important population parameters.

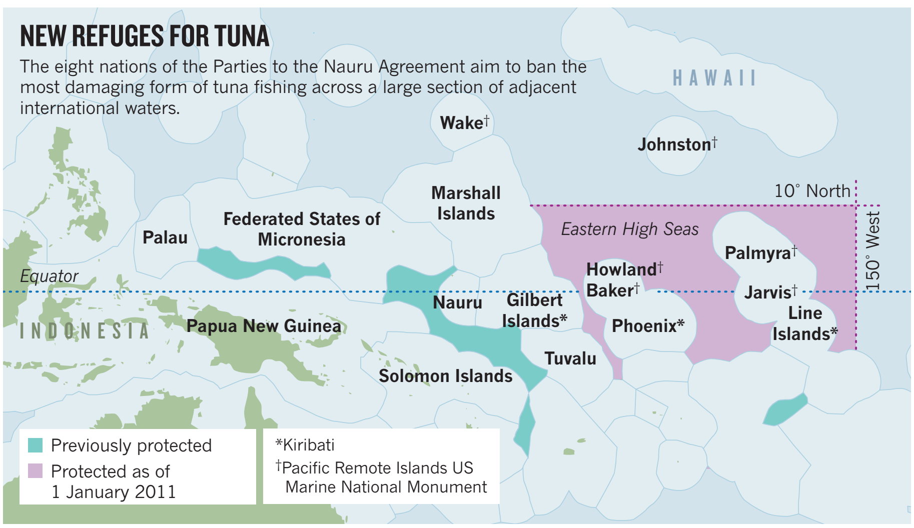
"These are the most far-reaching ocean-conservation measures ever."

"With good tagging data, we'll have a better understanding of the way the tuna move, and that will help us predict the effects of the closures on the population levels," he says.

Pauly predicts that because the individual tuna that do not travel long distances will survive in greater numbers, their offspring will have a genetic advantage over those that do range more widely, enhancing the conservation value of the refuges. "It's going to make the islands and the seamounts much more important, more attractive to these fishes," Pauly says.

Because this is the first time that anyone has tried to end fishing in a large body of international waters, it could establish a precedent, observers say. Until now, the world's seas have been fair game to the international industrial fishing fleets, with the commissions that theoretically have the power to restrict them strongly influenced by the fleets themselves.

"The closures would raise legal issues if they weren't justified as conservation measures," says Satya Nandan, chairman of the WCPFC. "Our stocks are relatively healthy and we want to keep them that way." ■



BIOCHEMISTRY

Microbe gets toxic response

Researchers question the science behind last week's revelation of arsenic-based life.

BY ALLA KATSNELSON

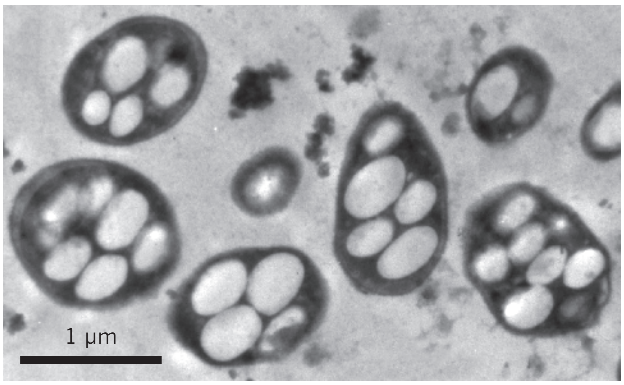
Days after an announcement that a strain of bacteria can apparently use arsenic in place of phosphorous to build its DNA and other biomolecules — an ability unknown in any other organism — some scientists are questioning the finding and taking issue with how it was communicated to non-specialists.

Many readily agree that the bacterium, described last week in *Science* and dubbed GFAJ-1 (F. Wolfe-Simon *et al*. *Science* doi: 10.1126/science.1197258; 2010), performs a remarkable feat by surviving high concentrations of arsenic in California's Mono Lake and in the laboratory. But data in the paper, they argue, suggest that it is just as likely that the microbe isn't using the arsenic, but instead is scavenging every possible phosphate molecule while fighting off arsenic toxicity. The claim at a NASA press briefing that the bacterium represents a new chemistry of life is at best premature, they say.

"It's a great story about adaptation, but it's not ET," says Gerald Joyce, a biochemist at the Scripps Research Institute in La Jolla, California.

At the press briefing, Steven Benner, a chemist at the Foundation for Applied Molecular Evolution in Gainesville, Florida, who was invited to the event to offer outside comment, used the analogy of a steel chain with a tinfoil link to illustrate that the arsenate ion said to replace phosphate in the bacterium's DNA forms bonds that are orders of magnitude less stable. Not only would the organism's DNA have to stay together in spite of the weaker bonds, says Benner, but so would all the molecules required to draw arsenate from the environment and build it into the genetic material. Co-authors of the paper, including Paul Davies, an astrobiologist at Arizona State University in Tempe, have countered that the arsenate bonds could be reinforced by specialized molecules, or that arsenic-based life simply has a higher turnover for molecular disintegration and assembly than does conventional life.

The big problem, however, is that the authors have shown that the organism takes up arsenic, but they "haven't unambiguously identified any arsenic-containing organic compounds", says Roger Summons, a biogeochemist at the Massachusetts Institute of Technology in Cambridge. "And it's not difficult to do," he adds, noting that the team could have directly confirmed or disproved the presence of arsenic in the DNA or RNA using targeted mass spectrometry.



SCIENCE/AAAS

The cells' large vacuoles may indicate that they are sequestering arsenic.

Some researchers suggest that the authors' own data hint at an organism that is simply absorbing and isolating arsenate while making use of the trace phosphates in its environment. For one thing, says Joyce, the paper shows that the organisms appear bloated, and contain large, vacuole-like structures — often a sign of sequestered toxic material. The arsenate-grown cells were analysed in their resting phase, which requires less phosphate for survival than does active growth, notes Joyce, and cells grown in high concentrations of arsenate did not seem to contain any RNA — possibly because RNA production had shut down to conserve phosphate. One calculation in the paper showed that the DNA in arsenate-grown cells actually contained 26 times more phosphorus than arsenic.

"I fault the authors for not noticing these things and sorting them out," says Rosemary Redfield, a microbiologist at the University of British Columbia in Vancouver, Canada, whose summary of the paper's problems, posted on her blog on 4 December, has already had more than 30,000 hits. "We shouldn't have to do the thinking for them."

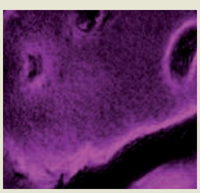
Felisa Wolfe-Simon, a NASA astrobiology research fellow at the US Geological Survey in Menlo Park, California, and the study's lead author, refused to address criticisms. "We are not going to engage in this sort of discussion," she wrote in an e-mail to *Nature*. "Any discourse will have to be peer-reviewed in the same manner as our paper was, and go through a vetting process so that all discussion is properly moderated."

But Jonathan Eisen, a microbiologist at the University of California, Davis, calls this "ludicrous", after a NASA press release drew media attention with claims of an "astrobiology finding that will impact the search for evidence of extraterrestrial life", a theme that Wolfe-Simon echoed at the briefing. "It is absurd for them to say that they are only going have the discussion in the scientific literature, when they started it," he says.

Ginger Pinholster, a spokeswoman for *Science*'s publisher, the American Association for the Advancement of Science in Washington DC, noted that the journal regards significant responses to high-visibility articles, as well as efforts to replicate the work, as a "key goal of publication". Pinholster also pointed out that the journal's own press summary of the paper made no mention of the search for extraterrestrial life, nor did *Science* "organize any additional promotional events". ■

MORE ONLINE

VIDEO STORY



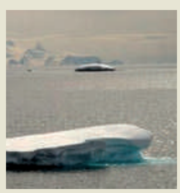
SCIENCE/AAAS

A microscope that makes movies of molecules in living animals go.nature.com/koghry

DAILY NEWS

- Space wing gets a lift from light go.nature.com/ye4zid
- What made swine flu worse? go.nature.com/cw3nrj
- Shocking behaviour in mercury nuclei go.nature.com/vxft1e
- Diversity is healthy go.nature.com/zdyaf9

EXPEDITION BLOG



J. QIU

Reporter Jane Qiu visits the Antarctic Peninsula, ground zero for climate change go.nature.com/wspgih

POLICY

Basel Declaration defends animal research

Dialogue with public is key to reducing opposition over the use of lab animals.

BY ALISON ABBOTT

Animal activists last summer set fire to the alpine holiday home of Daniel Vasella, then chief executive of pharmaceutical giant Novartis of Basel, Switzerland, in one of relatively few violent attacks on scientists working with animals in German-speaking countries.

But in the past few years these scientists have been feeling the pressure in other ways — from animal activists who have attempted to publicly shame them or have sent threatening e-mails, and from legislation that increasingly restricts the use of animals in basic research.

Now, in a bid to reverse that trend, more than 50 top scientists working in Germany and Switzerland have launched an education offensive. Meeting in Basel on 29 November, they drafted and signed a declaration pledging to be more open about their research, and to engage in more public dialogue.

"The public tends to have false perceptions about animal research, such as thinking they can always be replaced by alternative methods like cell culture," says Stefan Treue, director of the German Primate Center in Göttingen. Treue co-chaired the Basel meeting, called 'Research at a Crossroads', with molecular biologist Michael Hengartner, dean of science at the University of Zurich, Switzerland. Outreach activities, such as inviting the public into universities to talk to scientists about animal research, "will be helpful to both sides".

The Basel Declaration reiterates the legal and ethical requirements of the signatories to reduce the use of animals as far as possible (see 'Lab animals used in Germany'), and to keep their suffering to a minimum. But it also forcefully disputes recent efforts to declare animal use in basic research less acceptable than animal use in experiments that may yield practical benefits. In recent legal cases in the two countries, courts have interpreted national laws as forbidding basic research on primates.

In Bremen, Germany, the local government decided in 2007 not to renew the licence of neuroscientist Andreas Kreiter to work on macaques because the work was "too far from applications"[1]. This put a stop to his research recording from the animals' brains as they performed simple tasks. The ban still holds.

This echoed a case in Switzerland, where Kevan Martin, a director of the Institute of Neuroinformatics in Zurich, had to halt his research programme to map the functional microcircuitry of the brain of macaques in 2006 after Zurich's authorities declined to renew his licence for primate work[2]. The authorities, which also banned other local projects involving primates, said that Martin's work offended the dignity of the animals — which has been protected in the Swiss constitution since 2004 — and would not reap practical benefits for society in the near term.

Martin appealed the decision in Switzerland's supreme court, but in September last year the court upheld the decision.

"That was a shock for the community, and was one of the main motivations for holding our meeting," says Hengartner. "For the first time in Switzerland, the law was making a distinction between basic and applied research, and arguing that basic research with non-human primates is less valuable than applied research," he says. "But in biomedicine they are one and the same; applied research stands on the shoulders of basic research."

Scientists in Germany — which, unlike Switzerland, is a member of the European Union (EU) — had another reason to sign the declaration. In September, the EU approved a directive on animal use in research that must be translated into national law in the 27 EU member states within the next two years. "There are a lot of broad terms in the directive — like 'severe pain' — that could allow countries like Germany to choose wording which makes the legislation more restrictive than intended," says Treue. In addition, some of the directive's rules are unscientific, he says. For example, it includes special rules for cats and dogs.

> ***"The animal issue is never going to go away."***

Treue is particularly concerned that the directive bans some experiments outright — such as those involving severe pain, or using great apes — rather than allowing ethical committees to regulate them. However, early drafts of the EU directive were much worse, says Treue — one of many scientists who talked to EU parliamentarians to help to ensure that a ban on basic research using primates was removed[3].

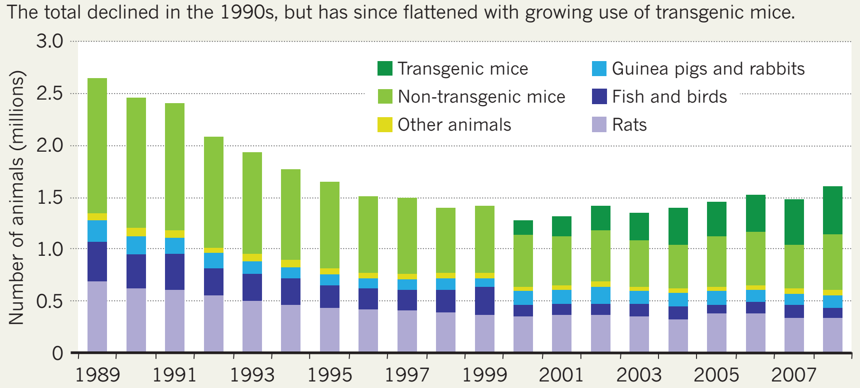
In the United Kingdom, similar initiatives coupled to animal terrorism laws helped to cut extremism, says Mark Matfield, director of the London-based European Biomedical Research Association, which represents animal researchers across Europe. "Being open and discussing with the public why you sometimes need to use animals is a reliable and tested idea which improved the climate for research in Britain."

The declaration will now be sent to deans of medicine and other research leaders in Germany and Switzerland to garner support. The organizers hope to go on to promote it internationally. "The animal issue is never going to go away," says Treue. "We need solidarity among all researchers." ■ SEE EDITORIAL, P.731

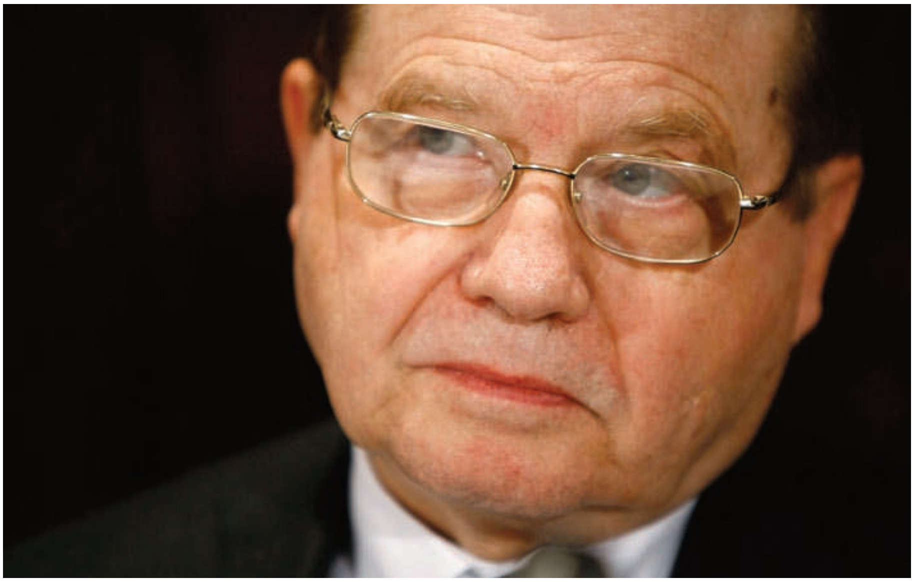
1. Schiermeier, Q. *Nature* **455,** 1159 (2008).
2. Abbott, A. *Nature* **453,** 833 (2008).
3. Abbott, A. *Nature* **464,** 964 (2010).

LAB ANIMALS USED IN GERMANY

The total declined in the 1990s, but has since flattened with growing use of transgenic mice.



SOURCE: UNIV. HEIDELBERG



W. MCNAMEE/GETTY

Luc Montagnier, co-discoverer of HIV, is no stranger to controversy.

MEDICINE

Trial draws fire

Nobel laureate to test link between autism and infection.

BY DECLAN BUTLER

Luc Montagnier is applying unorthodox ideas to the treatment of autism. With support from the Autism Research Institute (ARI), based in San Diego, California, the Nobel laureate is about to launch a small clinical trial of prolonged antibiotic treatment in children with autism disorders. The trial will also use techniques based on Montagnier's research into the notion that water can retain a 'memory' of long-vanished pathogens, and that DNA sequences produce water nanostructures that emit electromagnetic waves, published last year. But experts are critical and worry that the nobelist's status may lend unwarranted credibility to unconventional approaches to autism.

The planned pilot trial in France — funded by a US$40,000 grant from ARI — will screen around 30 children with autism disorders and 20 or so controls for bacterial infections, and then test whether months of antibiotic treatment improve the children's condition. Montagnier, who shared the 2008 Nobel prize for physiology or medicine for the discovery of HIV, concedes that there is no solid scientific evidence that infection causes or contributes to autism, but he argues that many parents and physicians have observed "spectacular" benefits from prolonged treatment. Stephen Edelson, director of ARI, says he's "very excited" about the "cutting-edge, groundbreaking" study.

Catherine Lord, a clinical psychologist working on autism at the University of Michigan in Ann Arbor, says that the trials are "not mainstream science". Lord says that many of the widely practised alternative medicine treatments for autism — including dietary modification, nutritional supplements and chelation therapy — are "semi-medical, not evidence-based science, and more pseudoscience."

Edelson, however, says that there are so many forms of autism and so much that is not known that "we need to study every angle". Criticisms of the science base of alternative approaches "probably would have been true were it ten years ago", he says, but critics don't appreciate how much research has been done since.

"I'm just interested in helping these children," Montagnier says. He acknowledges that many mainstream scientists are sceptical of his work, but defends his ideas. "In 1983, we were only a dozen or so people to believe that the virus we had isolated was the cause of AIDS."

Since then, Montagnier has supported non-mainstream theories in AIDS research that have put him at odds with other scientists. Most recently, he has argued that strengthening the immune system with antioxidants and nutritional supplements needs to be considered along with antiretroviral drugs in fighting AIDS, in particular in Africa.

NATURE.COM
For more on Nobel laureates' views on science see:
go.nature.com/fmmdne

"Montagnier's embrace of pseudoscientific and fringe agendas over the past few years has been seized on by AIDS denialists and other fringe groups, who make the case that Montagnier now supports their crazy views," says John Moore, an AIDS virologist at Cornell University in New York. Montagnier says that AIDS denialist groups misrepresent his thinking.

The autism trial enters a new area of controversy. The Infectious Disease Society of America have reviewed long-duration antibiotic treatments in Lyme disease, and concluded in April that the "inherent risks of long-term antibiotic therapy were not justified by clinical benefit". Montagnier acknowledges that safety concerns exist, but argues that opposition to long antibiotic treatments can also be "dogma". The trials will need to be cleared with the relevant ethics and regulatory bodies, he notes, and will include careful precautions and surveillance. "Expert physicians have learned to avoid side effects and to choose the right regimen," he says.

Another element of the trial is also attracting scepticism. Besides screening the children for pathogens with conventional DNA-amplification techniques, the researchers will use a diagnostic test based on the controversial idea championed by the late French scientist Jacques Benveniste, who claimed that water can retain the memory of substances it contained even after they have been diluted away. Studies have failed to confirm the claim, but Montagnier thinks that the 'memory' structures in the water can resonate with low-frequency electromagnetic signals, which he hopes can be transmitted over the Internet. He claims that very dilute solutions of pathogen DNA also emit such signals, and he intends to use this as a sensitive 'biomarker' for chronic infection.

Montagnier has published two papers on his research into the memory of water, one on bacterial DNA (ref. 1) and another claiming to have found electromagnetic signals of HIV DNA (ref. 2) in patients treated with antiretroviral drugs and whose blood seemed virus-free. He speculates that this may be the HIV reservoir from which the virus rebounds when antiretroviral treatment is paused or stopped. Several AIDS researchers contacted by *Nature* dismissed Montagnier's claims but declined to comment publicly.

"I'm just interested in helping these children."

Montagnier, who says that he is planning independent replication of his findings, published both papers in a new journal from Berlin-based Springer — *Interdisciplinary Sciences: Computational Life Sciences*, the editorial board of which he chairs. Asked why he didn't try to publish his astonishing findings in a higher-profile journal, Montagnier explained that he was sure that if he had sent them to *Nature* or *Science*, he would have run foul of experts who, on seeing mention of Benveniste or 'memory-of-water', would "reach for their revolvers". ■

1. Montagnier, L. *et al. Interdisciplin. Sci.: Comput. Life Sci.* **1,** 81–90 (2009).
2. Montagnier, L. *et al. Interdisciplin. Sci.: Comput. Life Sci.* **1,** 245–253 (2009).

B. EDWARDS/NEWSPIX/NEWS LTD

Restoration of Australia's drying Murray–Darling basin has caused a rift between farmers and scientists.

ECOLOGY

River chief resigns

Plan to save Australian river system runs aground.

BY STEPHEN PINCOCK IN SYDNEY

The challenge of balancing human needs for water with those of the environment has come to a head in Australia's ailing Murray–Darling river system, where an ambitious plan to restore the ecosystem was thrown into turmoil on 7 December after the leader of the scheme resigned.

Michael Taylor, chairman of the government-appointed Murray–Darling Basin Authority, announced he was stepping down just 2 months after releasing a draft plan for water management that prompted protests from the region's farmers, but which was largely backed by many of Australia's environmental scientists.

Covering 1 million square kilometres in southeastern Australia, the Murray–Darling basin is a vital agricultural region, containing rich wetland habitats and around 50 endangered species of birds and mammals. But decades of poor water management and drought have left ecosystems in crisis and many farmers short of water.

Research by the Commonwealth Scientific and Industrial Research Organisation in 2007–08 revealed that water use in the basin had reduced average annual streamflow at the mouth of the Murray River by 61%, which now fails to reach the sea 40% of the time. This has brought ecological problems such as increased salinity, algal blooms and the collapse of native fish and waterbird populations.

Successive state and national governments in Australia have struggled to manage the region's water resources. In October, the basin authority released a blueprint to address the problem by reducing the amount of water diverted for irrigation from the basin's 77,000 kilometres of rivers. The main recommendation — that 3,000–4,000 billion litres of water should be released from agricultural use each year and returned to the environment — triggered months of heated debate about the effect on farm communities.

The government asked Taylor to ensure the plan balanced ecological outcomes with the impact on jobs or other socioeconomic factors. But in his resignation statement, Taylor said he had received legal advice that the plan "cannot compromise the minimum level of water required to restore the system's environment on social or economic grounds". He urged the government to reconsider the plan's next phase.

Richard Kingsford, director of the Australian Wetlands and Rivers Centre at the University of New South Wales in Sydney, said that the resignation was "symptomatic of what a difficult process this is. It's true that a basin plan can't do everything."

Kingsford and 57 other environmental scientists recently released a statement saying that debate over the plan was too focused on short-term economic pain rather than the long-term economic benefits of a healthy river system. He and his colleagues argue that the next version of the basin plan, due early next year, should use 3,000–4,000 billion litres as the minimum release volume. "Michael Taylor's resignation might provide for a new form of leadership to come in and take the issue on," Kingsford says.

The final plan for the Murray–Darling basin is due to be tabled in the Australian parliament at the end of 2011 following consultation with the public, state governments and ministers. ■

ETHICS

Self-plagiarism case prompts calls for agencies to tighten rules

Technology is bringing down instances of duplication, despite variability in oversight.

BY EUGENIE SAMUEL REICH

Is plagiarism a sin if the duplicated material is one's own? Self-plagiarism may seem a smaller infraction than stealing another author's work, but the practice is under increasing scrutiny, as the eruption two weeks ago of a long-standing controversy at Queen's University in Kingston, Canada, makes clear.

Colleagues of Reginald Smith, an emeritus professor of mechanical and materials engineering at Queen's, say that up to 20 of Smith's papers contain material copied without acknowledgment from previous publications. University officials first learned of the duplications in 2005, and they eventually led to an investigation by the Natural Sciences and Engineering Research Council (NSERC), which funded some of Smith's work, including experiments on board the US space shuttles. Although Smith avoided censure for research misconduct, three papers were subsequently retracted by the *Annals of the New York Academy of Sciences*[1] and one by the *Journal of Materials Processing Technology*[2]. The situation was recently made public in news reports and has led to calls for stronger powers by funding agencies in Canada to discipline researchers who engage in the practice.

"He was a very good scientist, but something happened and he got into this business of duplicating papers," says Chris Pickles, a metallurgist at Queen's who raised concerns about Smith's publication practices after spotting some duplications under Smith's name while searching an online database. Smith referred a request for comment to his lawyer, Ken Clark of law firm Aird and Berlis in Toronto, Canada, who notes that many of the republications duplicated material from conference proceedings, which in an earlier epoch would not usually have been published. He also notes that Smith is retired, and does not stand to gain financially from his republications.

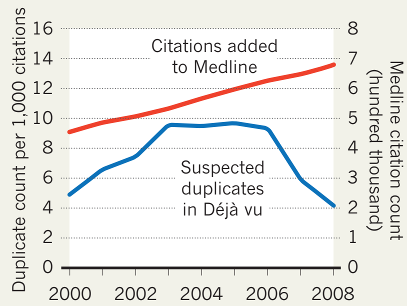
Many researchers say that republication without citation violates the premise that each scientific paper should be an original contribution. It can also serve to falsely inflate a researcher's CV by suggesting a higher level of productivity. And although the repetition of the methods section of a paper is not necessarily considered inappropriate by the scientific community, "we would expect that results, discussion and the abstract present novel results", says Harold Garner, a bioinformatician at Virginia Polytechnic Institute and State University in Blacksburg. Garner's research group used an automated software tool to check the biomedical literature for duplicated text, and identified more than 79,000 pairs of article abstracts and titles containing duplicated wording. He says work on the database of partly duplicated articles — called Déjà vu (go.nature.com/hgq2t4) — has led to close to 100 retractions by journal editors who found the reuse improper. An analysis by Garner in the press at *Urologic Oncology*[3] shows that while the total quantity of biomedical literature has risen steadily since 2000, cases of republication stopped rising after 2003 and fell sharply between 2006 and 2008 (see graph). "It actually does look like it's getting better," says Garner. "People who would ordinarily step across the line are not doing it."

He credits increased vigilance by journal editors who are using his free tool or commercially available software to check submissions for repeated text and halt dubious papers before they reach publication.

> ➲ **NATURE.COM**
> Journals step up plagiarism policing:
> go.nature.com/kdmlsa

DROP IN DUPLICITY?

There has been a decline in the number of new highly similar pairs of manuscripts.



SOURCE: ADAPTED FROM REF 3; ELSEVIER

NSERC's policy on integrity in research makes no specific reference to plagiarism or self-plagiarism, which has led to calls for tougher rules in the wake of the publicity over Smith's case. In the United States, the National Science Foundation (NSF) takes a strong stance on plagiarism in general, says Christine Boesz, who was inspector-general at the NSF from 1999 until 2008. "The NSF got into the plagiarism game early," she says. Numbers obtained by *Nature* under the US Freedom of Information Act show that, since 2007, the agency has found between 5 and 13 cases of plagiarism each year. In contrast, the US Department of Health and Human Services's Office of Research Integrity (ORI), which is responsible for overseeing alleged plagiarism associated with National Institutes of Health research, has reported no cases of plagiarism of text over the past three years, but has found up to 14 scientists a year guilty of falsification or fabrication of data (see table).

Ann Bradley, a spokeswoman for the ORI, says the office's working definition of plagiarism (go.nature.com/p15kcu) excludes minor cases. Nick Steneck, director of research ethics and integrity at the University of Michigan in Ann Arbor, says authorities worldwide should adopt a uniform misconduct policy that provides clear guidance not only on data falsification and fabrication but also on lesser ethical breaches — such as self-plagiarism ■

1. Braaten, D. *Ann. NY Acad. Sci.* **1176,** 228 (2009).
2. Smith, R. W., DeMonte, A. & Mackay, W. B. F. *J. Mater. Process. Tech.* **153–154,** 589–595 (2004).
3. Garner, H. R. *Urol. Oncol.-Semin. Ori.* doi:10.1016/j.urolonc.2010.09.016 (in the press).

CASES OF MISCONDUCT AND PLAGIARISM AS REPORTED BY US RESEARCH AGENCIES

	Office of Research Integrity (ORI)			National Science Foundation (NSF)		
	2008	2009	2010	2008	2009	2010
Debarments for falsification/fabrication	2	1	3	1	2	3
Debarments for plagiarism	0	0	0	0	4	2
Findings of falsification/fabrication	7	14	7	1	1	2
Findings of plagiarism	0	0	0	5	13	10

Number of funded researchers: National Institutes of Health (ORI) 325,000; NSF 98,820 (2010). 2010 data run until August.

SOURCES: ORI; NSF

MOLECULAR BIOLOGY

Stem-cell progress

Replacing genes with drugs could allow safe reprogramming.

BY EWEN CALLAWAY

Ever since scientists first switched adult human cells into an embryonic-like state from which they can develop into any tissue type, recipes for making these induced pluripotent stem (iPS) cells have multiplied.

However, many rely on the introduction of foreign genes by viruses, which makes the altered cells unsuitable for use in patients. Now researchers have replaced all but one of the genes with a cocktail of chemicals, taking scientists a step closer to creating patient-specific iPS cells that could be used in the clinic.

The advance, by chemist Sheng Ding at the Scripps Research Institute in San Diego, California, and his colleagues (Zhu, S. *et al. Cell Stem Cell* **7,** 651–655; 2010) is an adaptation of an approach developed by Shinya Yamanaka's lab at Kyoto University in Japan. Yamanaka's group — one of two to first create iPS cells — infected adult cells with viruses carrying the genes *OCT3/4*, *SOX2*, *KLF4* and *c-MYC*.

However, these iPS cells have foreign DNA peppered throughout their genomes, where it might interrupt genes that protect against tumours. Over the past few years, scientists such as Ding have developed safer ways to make pluripotent cells, by delivering reprogramming factors in other ways (see 'Virus-free iPS cells').

In 2008, Ding's team showed that a mixture of chemicals and two genes could reprogram neural progenitor cells, which already express other genes needed to make iPS cells (Shi, Y. *et al. Cell Stem Cell* **2,** 525–528; 2008). Now his group has made human iPS cells from skin cells by treating them with drugs and just one virus-delivered gene, *OCT4*. The resulting cells express the same genes as embryonic stem cells and can transform into different types of cell.

OCT4 can be replaced as well, so an iPS protocol entirely free of foreign genes shouldn't be far off. Ding says that his team has already created mouse iPS cells using only drugs, and is making progress with human cells.

Robert Lanza, chief scientific officer of Advanced Cell Technology in Marlborough, Massachusetts, says that iPS cells should soon be safe enough to test in humans. "I think we now have the tools to contemplate clinical trials." ■

VIRUS-FREE IPS CELLS

Reprogramming-factor delivery method	Pros	Cons
Proteins	No foreign genes involved	Protein modifications could influence reprogramming, expensive to manufacture
Transposons	Just one piece of DNA is inserted into the genome	The transposon must be removed using a specialized enzyme
RNA	Quicker and more efficient than other methods	Potential for immune response, expensive to manufacture
Small molecules	No foreign genes, potentially cheap to make	Drugs can affect unintended protein targets

THE SUPERSOLID'S NEMESIS

John Reppy has come out of retirement to question the high-profile discovery of a new kind of quantum matter.

BY EUGENIE SAMUEL REICH

S. OGDEN

The fourth time he is asked what the dental floss is for, John Reppy seems to hear. He picks up a pair of scissors, and starts snipping away at the plastic strands wound round the shiny beryllium–copper components of his torsional-oscillator experiment. "I want to make a change to it anyway," he says. As he snips, pieces of wire and piping begin to pop out of the neat cylindrical column he has built, making it completely clear what the floss is for: to hold everything down.

The pieces of this experiment, in a basement lab at Cornell University's Clark Hall in Ithaca, New York, span more than half of Reppy's 50-year career studying the behaviour of helium cooled to ultra-low temperatures. Near the top of the metre-long column is a 30-year-old refrigeration unit that Reppy found among the bric-a-brac in his lab a few years after he signed up for retirement. Below that is a torsional oscillator of the type he invented in the 1970s — a cylindrical vessel, just a few centimetres across, that is free to twist back and forth around a rod running down the centre of the cylinder. When the vessel is filled with the isotope helium-4 via pipes wrapped around the column, and when its temperature is gradually lowered, changes in the oscillation frequency reveal changes in the physical properties of the helium. At two-tenths of a degree above absolute

John Reppy holds up a torsional oscillator used in his experiments.

zero, for example, the helium-4 condenses into a solid crystal — and may even turn into a 'supersolid', a strange quantum state in which some of the atoms seem to pass through others without friction.

Or it may not. In recent months, the results from his apparatus have led the 79-year-old, semi-retired Reppy to become a vocal critic of a 2004 claim by physicists Moses Chan and Eun-Seong Kim that they had formed a supersolid in Chan's laboratory at Pennsylvania State University in University Park. The stakes are high: other such macroscopic-scale quantum effects, such as superconductivity and superfluidity, have won their discoverers Nobel prizes. And Reppy knows his criticisms are raising hackles in the field. Others have replicated Chan and Kim's results, yet no one has replicated Reppy's contradictory finding. "My result is a bolt out of the blue," he admits.

Nonetheless, as the inventor of the modern torsional-oscillator apparatus, and as Chan's former supervisor, Reppy has a professional stature that makes his views impossible to ignore. "He's come up with a lot of inventive, clever experimental techniques, and always manages to pick out the experiment that reveals what's really going on," says David Lee at Texas A&M University in College Station, one of the winners of the 1996 physics Nobel prize for the discovery of superfluidity in another isotope of helium, helium-3, work done while he was at Cornell.

HEART OF THE MATTER

The roots of the supersolid controversy go back to 1969, when Russian physicists predicted a state of solid matter in which gaps, or vacancies, in a crystal structure could move together as a single quantum wave — a collective motion reminiscent of the frictionless flow of a superfluid.

In 2004, Chan and Kim reported the first experimental evidence consistent with this 'supersolid' behaviour[1,2]. They found that the back-and-forth swings of a torsional oscillator filled with solid helium-4 sped up as the temperature was lowered to below two-tenths of a degree above absolute zero — just as would be expected if a supersolid were forming inside. The idea is that the zero-friction quantum effects predicted by the Russians effectively decouple some of the atoms in the solid and prevent them from oscillating along with the rest of the atoms. This makes the inertia of the oscillator smaller than the total quantity of helium would suggest, which leads to the faster oscillations. Chan and Kim's claim prompted enormous excitement, and about a dozen researchers began building torsional oscillators in a bid to replicate the observation.

Reppy was one of them. When Chan and Kim first reported their results, Reppy was approaching the end of a five-year Cornell programme intended to ease older faculty members into retirement by steadily reducing their hours, teaching responsibilities and lab space. He was looking forward to a retirement spent rock-climbing — a field in which his reputation looms as large as it does in physics. A world-class climber since his student days, Reppy is famous for his invention and promotion of clean-climbing techniques, in which the nuts that hold the rope are wedged into existing, natural cracks in rock faces rather than banged in like pitons. He says he likes the technique not just because it is environmentally friendly, but because it is easy. And safe: when he talks about climbing, he doesn't emphasize the obvious excitement it gives him, so much as his caution. "You always climb with a partner," he says.

But with supersolidity promising a different kind of adventure, Reppy decided to make a comeback from retirement. When he heard that a fellow faculty member at Cornell had taken on a graduate student, Sophie Rittner, to replicate Chan and Kim's experiment, Reppy suggested that she work with him instead: just as in climbing, he needed a partner. Rittner did not immediately jump at the idea of signing on with a retired professor with a lab full of junk and no active research group. But it was obvious that Reppy would be a good supervisor for her work, as he had had a long career developing experimental tricks for studying superfluidity in helium-3. "I came to appreciate the fact that I had an adviser with a huge amount of time. He was super hands-on," says Rittner. Together they ordered the parts for a torsional oscillator, which Rittner constructed. Then every morning, Reppy came into the lab at about 7 a.m. and started the experiment going. He would stay until about 4 p.m.; Rittner came in later in the morning and stayed into the evening.

The pair soon saw the increased oscillation that Chan and Kim had reported. But in February 2006, they tried something new and got a surprise. After one run of recording data, Reppy and Rittner let the frozen helium warm up to just above 1 kelvin, and then lowered the temperature again to repeat the run. The second time around, the speed-up was markedly diminished.

Heating and then recooling a crystal, a process called annealing, is in general expected to remove defects in the crystal structure. To Reppy and Rittner, the implications of their observations were clear: the supersolid signal was not due to an intrinsic quantum behaviour of a pure crystal, but was somehow caused by disorder in the structure, which is why it went away when the defects did.

IN A SPIN

When Rittner presented the results[3] at the March 2006 American Physical Society meeting, there was something of an uproar, she says. "People were saying 'what is this?'" The findings threatened to make supersolidity substantially less interesting, because effects caused by imperfections and impurities often turn out to be impossible for theoretical physicists to calculate exactly. Even now, six years after Chan and Kim's experiment was published, there is still no comprehensive theory of supersolidity. And when Rittner gained her PhD, she decided to move on to a different research field.

Left on his own, Reppy — an inveterate tinkerer — was soon trying to improve his apparatus. Picking up a box containing many of his historical torsional oscillators, he gives it a rattle, selects one and points out an interesting ridge. He loves to shape the metal pieces himself, he says. And with no more administrative duties to distract him, he adds happily: "I can spend all my time in the machine shop."

One of Reppy's first moves after Rittner left was to make a new oscillator vessel, which, instead of holding the helium in a ring-shaped channel

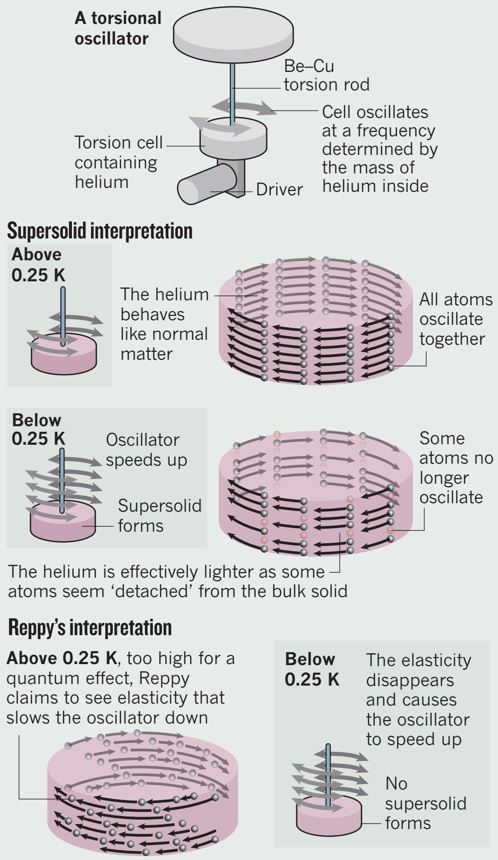
S. OGDEN

Craftsman at work: Reppy loves to fashion the parts he needs for his research.

circling the cylinder's rim, also included a channel across the middle. He filled the oscillator with helium-4, ran the apparatus and verified that he saw the faster oscillations attributed to a supersolid. Then he blocked one side of the ring, so that the putative supersolid could flow only through the central channel and the other side of the ring, and found that the signal decreased — just as the supersolid theory would predict. But then he also blocked the central channel to try to stop the supersolid flowing at all, which should have made the signal go away entirely. But it didn't. Thinking that there must be a leak, Reppy tried several variations — including just watching to see if the helium-4 escaped like air out of a balloon. It didn't.

Reppy has never understood this observation, and hasn't published it. But the unexpected behaviour planted a seed of doubt in his mind: was the formation of a supersolid the true explanation for the effects that everyone had seen? There were other discordant findings too. For example, liquid helium-4 ought to be able to flow through a solid helium-4 barrier if that solid contains some supersolid. But neither Reppy and Rittner, nor a group led by John Beamish at the University of Alberta in Edmonton, were able to observe this.

By late 2009, Reppy had tinkered with his apparatus yet again, adding a diaphragm on top that allowed him to deform a sample during a measurement run. Following up on the possibility that disorder was involved in supersolidity, he wanted to see if he could increase the amount of supersolidity in a given sample by using the deformation to introduce more defects. The results of this experiment were totally unexpected: Reppy found no evidence of a supersolid signal at all — at least, not at ultra-low temperatures. Instead, the deformation produced a decrease in the oscillation frequency at higher temperatures — so high that the jiggling of atoms would be expected to destroy any quantum effect such as supersolidity.



SUPERSOLID OR NOT?
Experts disagree on whether helium enters a rare quantum state at very low temperatures

SOLID GROUND?

The publication of these results[4] in June 2010 caused another stir, including a news article in *Science* claiming that the evidence for a supersolid was "slipping away"[5]. In a commentary in *Physics*[6], Beamish suggested that Reppy had discovered a kind of "quantum plasticity", an effect in which solid helium-4 radically increases its softness as its temperature is raised, then stiffens again as the temperature is lowered. That stiffening would cause the frequency of a torsional oscillator to increase and mimic the supersolid signal. Reppy has embraced that idea and interpreted his results as a repudiation of supersolidity — an indication that he, Chan, Kim and everyone else had in fact been seeing the disappearance of the previously undetected quantum plasticity at low temperatures. He insists that he takes no joy in that conclusion. "I'm disappointed that this is turning out to be something other than a supersolid," he says.

But others in the community are not so convinced. Is the evidence for supersolidity really slipping away, or does Reppy just have anomalous equipment? Sebastien Balibar, an expert on helium-4 at the École Normale Supérieure in Paris, says he believes two novel effects have been discovered — supersolidity and a radical change in elasticity, something akin to Beamish's quantum plasticity. Without meaning to, Balibar says, Reppy configured his vessel to be especially sensitive to changes in the elastic properties of the material — perhaps because the stiffer helium-4 at lower temperatures was effectively gluing together parts of his experimental chamber, and producing a heavier oscillator.

Asked about that possibility, Reppy whisks the visitor out of his lab to a blackboard in a breezy corridor, where he chalks out a calculation showing why he feels the gluing hypothesis is extremely unlikely.

Even so, Reppy is having difficulty getting others in the field to share his doubts about the supersolid. Kim and his colleagues have just published additional results[7] showing what they say is conclusive evidence for formation of a supersolid. Reppy has seen Kim's work, and says that he feels the problem presented by his own results hasn't yet been addressed properly. But the field of supersolid helium-4 is too small and collaborative for Reppy's result to be ignored. Chan — for one — naturally bridles at the suggestion that the supersolid interpretation is in trouble, but he has also asked Reppy to collaborate, to gain a better understanding of his equipment. Chan points out that even when Reppy and Rittner replicated the 2004 experiment, they were reporting supersolid fractions of 20% — 20 times greater than the 1–2% measured by other groups. He takes that as evidence that there's a secondary effect at work in Reppy's apparatus that is swamping the supersolid signal. Chan hopes that could be understood by testing the vessel's response when filled with better-studied superfluid helium-3.

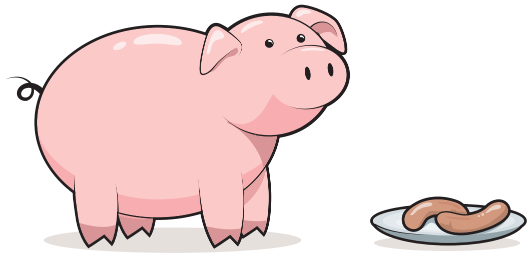
Meanwhile, Reppy's latest, unpublished, results are giving him new cause for doubt. These data were taken with a secondary oscillator added to the bottom of his experiment, which allows him to vary the frequency of the vessel's oscillation. His preliminary finding is that the response of the helium-4 sample depends on that frequency — which would not be the case if the helium-4 was a supersolid. But, Reppy wonders, doodling with chalk on a cartoon sketch of his vessel, could this be a way to turn the critiques of his experiment into a bonus? He starts drawing an alternative configuration of the apparatus, in which he could produce the first measurement of the elasticity of helium-4. Asked what light that would shed on the formation — or otherwise — of a supersolid, he shrugs. "I don't know," he says.

Chan says that in a similar situation, with an experiment giving very surprising results, he probably wouldn't have published anything. But researchers in this field are having to feel their way experimentally because of the absence of a guiding theory. And, as tends to happen with a quintessential experimentalist, Reppy's caution inevitably gives way to dogged determination once he is confident that each result is real. "That 20% — he knows it's unusual, but he felt compelled to publish it," Chan says. "Whatever way it turns out, I think respect for him will grow." ■

Eugenie Samuel Reich *is a reporter for Nature.*

1. Kim, E. & Chan, M. H. W. *Nature* **427,** 225–227 (2004).
2. Kim, E. & Chan, M. H. W. *Science* **305,** 1941–1944 (2004).
3. Rittner, A. S. C. & Reppy, J. D. *Phys. Rev. Lett.* **97,** 165301 (2006).
4. Reppy, J. D. *Phys. Rev. Lett.* **104,** 255301 (2010).
5. Cho, A. *Science* **329,** 20 (2010).
6. Beamish, J. *Physics* **3,** 51 (2010).
7. Choi, H., Takahashi, D., Kono, K. & Kim, E. *Science* doi:10.1126/science.1196409 (2010).

➲ NATURE.COM
For more on supersolids see:
go.nature.com/hbwgp7



A taste of things to come?

Researchers are sure that they can put lab-grown meat on the menu — if they can just get cultured muscle cells to bulk up.

BY NICOLA JONES

Mark Post has never been tempted to taste the 'fake' pork that he grows in his lab. As far as he knows, the only person who has swallowed a strip of the pale, limp muscle tissue is a Russian TV journalist who visited the lab this year to film its work. "He just took it with tweezers out of the culture dish and stuffed it in his mouth before I could say anything," says Post. "He said it was chewy and tasteless."

Post, who works at the Eindhoven University of Technology in the Netherlands, is at the leading edge of efforts to make *in vitro* meat by growing animal muscle cells in a dish. His ultimate goal is to help rid the world of the wasteful production of farm animals for food by helping to develop life-like steaks. In the near term, he hopes to make a single palatable sausage of ground pork, showcased next to the living pig that donated its starter cells — if he can secure funds for his research.

Post started out as a tissue engineer interested in turning stem cells into human muscle for use in reconstructive surgery, but switched to meat a few years ago. "I realized this could have much greater impact than any of the medical work I'd been doing over 20 years — in terms of environmental benefits, health benefits, benefits against world starvation," he says. Largely because of the inefficiency of growing crops to feed livestock, a vegetarian diet requires only 35% as much water and 40% as much energy as that of a meat-eater[1]. Future 'in-vitrotarians' should be able to claim similar savings.

The prospect of an alternative to slaughtering animals led People for the Ethical Treatment of Animals based in Norfolk, Virginia, to announce two years ago a US$1-million prize for the first company to bring synthetic chicken meat to stores in at least six US states by 2016. In the Netherlands, where the vast majority of work has been done so far, a consortium of researchers convinced the government to grant them €2 million (US$2.6 million) between 2005 and 2009 for developing *in vitro* meat.

Such incentives have helped to solve some of the basic challenges, applying human tissue-engineering techniques to isolate adult stem cells from muscle, amplify them in culture and fuse them into centimetre-long strips. But far more money and momentum will be needed to make *in vitro* meat efficient to produce, cheap and supermarket-friendly. Post estimates that creating his single sausage will require another year of research and at least $250,000. So what still needs to be done?

CHOOSE THE RIGHT STOCK

The first question for researchers is which cells to start with. Embryonic stem cells would provide an immortal (and therefore cheap) stock from which to grow endless supplies of meat. But attempts to produce embryonic stem cells from farm animals have not been successful. Most work so far has been on myosatellite cells, the adult stem cells that are responsible for muscle growth and repair. These can be obtained by a relatively harmless muscle biopsy from a pig, cow, sheep, chicken or turkey; the desired cells are then extracted using enzymes or pipetting, and multiplied in culture.

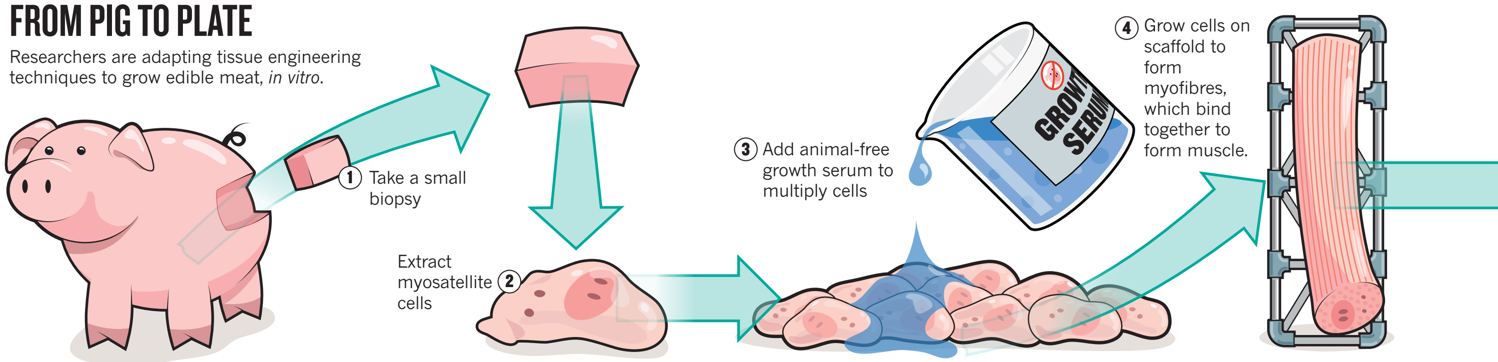
Morris Benjaminson, professor emeritus at Touro College in New York, prefers a different approach — planting the whole biopsy in a dish. "We use the whole business without the brain," he says. "We don't break it down to cells and put them back together again." He used this method to grow goldfish fillets in his lab in 2002, boosting the surface area by up to 79% over a week by adding a shot of extra cells from ground-up muscle[2]. It's not clear, however, whether this procedure could produce enough muscle for a commercial enterprise.

The fundamental problem is that myosatellite cells will only divide dozens of times, probably because their telomeres — the protective ends of the chromosomes — wear down with age. There are ways of boosting their proliferation. One is to add a gene for the repair enzyme telomerase. Another, being investigated by the start-up company Mokshagundam Biotechnologies in Palo Alto, California, involves inserting a tumour-growth-promoting gene. But genetically modified lab-grown meat might be too much for consumers to swallow. "Try selling that," laughs Post. An alternative is to get cells from a young

ILLUSTRATIONS: NIK SPENCER

FROM PIG TO PLATE

Researchers are adapting tissue engineering techniques to grow edible meat, *in vitro*.



animal and perfect the rest of the system, such as the culture medium, to maximize growth.

For now, Post uses regular cell-culture medium to grow his pork myosatellite cells. This contains fetal calf serum which, as it currently comes from dead cows, largely defeats the point of synthetic meat. It also contains antibiotics and anti-fungal agents that might not be good for human consumption. "Supposedly you could be allergic to these; you never know," Post says. To get the cells to differentiate into muscle, he shifts to horse serum, which has the same list of problems.

Animal-free media made from a slurry of plants or microbes are commercially available for biomedical work such as *in vitro* fertilization. But like animal-based media, they're expensive — right now, growth media account for about 90% of the material costs of lab-grown meat. And their composition is proprietary, making them difficult to customize. One alternative might be to use ground-up maitake mushrooms, which Benjaminson found works just as well as calf serum for his fish fillets. At the University of Amsterdam, researchers working on *in vitro* meat have been developing a cheap medium made from blue-green algae, with added growth factors made in genetically modified *Escherichia coli*. But no one has yet developed a way of making a cheap, animal-free growth serum in large quantities.

BEEF IT UP

Myosatellite cells grown on a scaffold will fuse into myofibres, which then bundle together to make up muscle. But lab-assembled muscles are weak and textureless. "It's like when you take off a cast after six weeks," says Post. To get the muscle to bulk up with protein requires exercise. Assembling the myofibres between anchor points helps, as this creates a natural tension for the muscle to flex against. Post uses this type of arrangement to boost the protein content of a muscle strip from 100 milligrams to about 800 milligrams over a few weeks. He also administers 10-volt shocks every second, which can bump protein content up to about a gram. This much electricity would be expensive in a scaled-up industrial process, so his group is hoping to learn how to mimic chemical signals that tell muscles to contract.

Vladimir Mironov of the Medical University of South Carolina in Charleston is instead using a scaffold made of chitosan microbeads — chitosan can be sourced from crabs or fungi — that expand and contract with temperature swings, thus making a natural fitness centre for his muscle strips.

If lab-grown muscle gets more than about 200 micrometres thick, cells in the interior start to die as they become starved of nutrients and oxygen. Post simply grows many small strips that could be ground up into a sausage. Others, including Mironov, are using blender-sized bioreactors of the type developed by NASA to study muscle growth in low gravity. These conditions help prevent cell clumping and improve transport of oxygen and nutrients.

Growing meat on an industrial scale would require large, customized bioreactors like those used by biopharmaceutical companies. Mironov estimates that a commercial *in vitro* meat facility would need a five-storey building of bioreactors; with a similarly huge investment. And all that is just for manufacturing ground meat. The prospect of growing steaks is a much bigger challenge, requiring a system of fake 'blood vessels' built into the meat. That is decades away.

"This could have much greater impact than any of the medical work I'd been doing over 20 years."

MARKET IT

The thing that enthusiasts for fake meat talk least about is its taste, perhaps because they haven't tried it. In the United States, researchers have largely avoided eating anything grown in the lab for fear of violating a Food and Drug Administration regulation (it's unclear whether it is actually forbidden) or being seen as publicity hounds. When Benjaminson grew his goldfish fillets, his team dipped them in olive oil, fried them in breadcrumbs and gave them to an 'odour and sight' panel who said they seemed edible, but who weren't allowed to try them.

Researchers generally believe that if they can get the texture right, taste will follow — particularly once flavouring is added. Fortunately, myosatellite cells can also turn into fat, which would add to the taste. At Mokshagundam Biotechnologies, the goal is to make a spam-like mix of different muscle and other cell types that provide the 'umami' taste that characterizes meat. Scientists will also have to find a way of adding nutrients such as iron (which comes from blood) and vitamin B12 (which comes from gut bacteria).

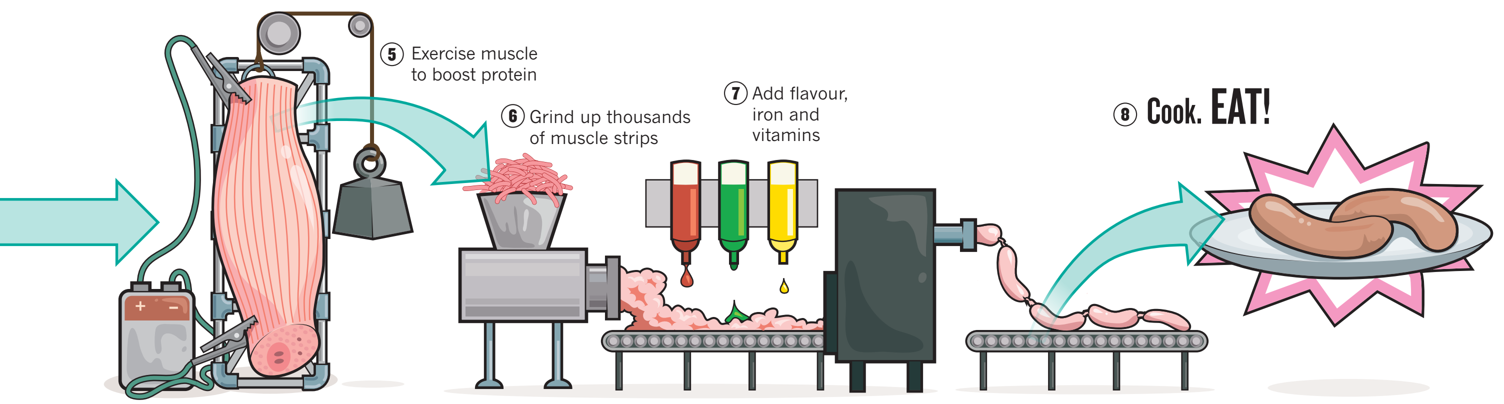
The process won't be cheap. By one rough estimate, first-generation lab meat could cost €3,500 per tonne (compared with €1,800 per tonne for unsubsidized farmed chicken meat)[3]. Mironov thinks the best way to secure an early market is by turning *in vitro* meat into a 'functional' food attractive to the rich and famous, perhaps by filling it with compounds that promote health or suppress appetite. "Only Hollywood celebrities like Paris will be eating this," he says. Alternatively, one could get an edge on the market by making meat products from exotic or even extinct animals, assuming a few of their cells could be saved. In the long run, advocates see a market in vegetarians and others who want guilt-free and environmentally friendly meat.

Researchers such as Post believe that the scientific and technical advances needed to make and sell *in vitro* meat are worth the fight — but convincing funders remains the biggest obstacle. Post's funding from the Dutch government ran out in 2009; he came out of that with hundreds of pork strips, not the thousands he needs for a sausage. Today, a couple of umbrella organizations promote the cause, including the non-profit New Harvest, which provides small funds for US- and Europe-based work, and a new commercial company, California-based Pure Bioengineering, which aims to raise venture capital. But no windfalls have arrived as yet.

Post will keep seeking funding for his demonstration sausage — but he knows that raising enough to commercialize the entire process will be a huge ask. "I usually say €100 million," says Post. "That's the number I forward to the government, and then they faint." ■

Nicola Jones *is a freelance journalist based in Vancouver, Canada.*

1. Marlow, H. J. *et al. Am. J. Clin. Nutr.* **89,** 1699S–1703S (2009).
2. Benjaminson, M. A., Gilchriest, J. A. & Lorenz M. *Acta Astronaut.* **51,** 879–889 (2002).
3. eXmoor pharma concepts. *The* in vitro *Meat Consortium Preliminary Economics Study Project 29071* (2008); available at go.nature.com/firfka



P. MACDIARMID/GETTY

N. BHOJANI/BLOOMBERG NEWS/GETTY

Increased patenting and licensing could lead to pharmaceutical breakthroughs in developing countries such as India, but slow progress in other areas.

Lessons from Bayh–Dole

Developing countries wanting to boost commercialization of their academic research should learn from the mistakes of US patenting legislation, says **Bhaven N. Sampat**.

Thirty years ago this month, the US Congress passed the Bayh–Dole Act. The policy replaced a mishmash of rules that had governed the ownership of patents resulting from publicly funded research. Under the act, grantees and contractors, instead of government funding agencies, hold title to inventions.

Bayh–Dole has been widely celebrated for its effect on US universities. Since its passage, the number of patents that universities have been granted has climbed from fewer than 300 a year to more than 3,000. And, having earned very little from licensing before the act, US universities now earn almost US$2 billion annually[1].

Policy-makers in other countries have taken these trends as evidence that promoting patents and exclusive licensing on the outputs of taxpayer-funded research enhances technology transfer, commercialization and innovation. This has led numerous developing countries — including South Africa, the Philippines and Brazil — to enact Bayh–Dole-style legislation. Others, including India, are considering similar approaches.

Yet countries looking to boost commercialization should be wary of the myth that the act transformed US universities into entrepreneurial institutions capable of generating successful spin-off firms, high-tech jobs and self-sustaining research funds — and all at no cost to the taxpayer. Instead, they should note the problems that have arisen with the act, such as the overly restrictive patenting and licensing mentality it has generated among many technology-transfer offices, and craft their own legislation to avoid these pitfalls.

NATURE.COM
Scientists opt for consultancy over invention:
go.nature.com/kzra8k

The Bayh–Dole legislation was passed in response to a particular set of US problems at a particular time. An important motivation was to give universities the right to patent drug compounds, and to exclusively license them to companies. Before the act, to do either was difficult because of bureaucracy, particularly at the Department of Health, Education and Welfare. Policy-makers were also concerned that aggressive patent policies established in the 1960s by the National Institutes of Health's medicinal-chemistry ▶

programme had reduced collaboration between universities and industry.

Another major concern in the 1970s was the allegedly low rate of commercialization of federally funded research, including that conducted outside universities. Less than 5% of the 28,000 patents owned in 1976 by the government were licensed to industry[2].

The economic argument for allowing companies exclusive access to drug compounds is a strong one. Universities generated nearly one-fifth of the drugs with the greatest clinical impact approved by the US Food and Drug Administration during the past three decades[3]. It is hard to imagine that the profit-oriented companies who developed these drug candidates and put them through clinical trials would have invested the hundreds of millions of dollars needed if competitors could copy and market the drug themselves.

ANTIQUATED ARGUMENTS

Thirty years on, the 1976 licensing figure and the rise of university licensing since 1980 (see 'Technology transfer') form the central arguments used to claim that the Bayh–Dole Act was needed to boost technology transfer for all government-funded research, not just for pharmaceuticals. But these figures are misleading because they downplay the other ways in which universities contribute to economic growth and innovation. Researchers also disseminate their findings and ideas through consulting, publishing and teaching[4]. Indeed, the development of numerous US industries — including chemical engineering, aeronautics, computing and agriculture — relied heavily on academic research, but with little or no university patenting[5].

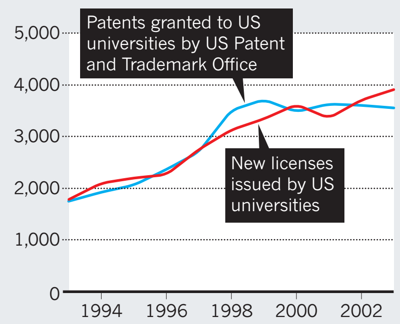
Although universities would probably not have made as much money, many of the non-drug technologies licensed after Bayh–Dole, including some of the most lucrative biotechnology techniques, would have been picked up anyway from academic publications and other traditional channels of dissemination. The Cohen–Boyer patent, for example, which covers recombinant DNA cloning techniques and is held jointly by Stanford University in California and the University of California, has generated more than $250 million, but even Niels Reimers, who managed Stanford University's licensing programme at the time, noted in a 1997 interview that "whether we licensed it or not, commercialization of recombinant DNA was going forward"[6].

In short, Bayh–Dole replaced one set of frictions with another — it eliminatated restrictions on patenting and technology-transfer licensing in favour of promoting excessive patenting and overly restrictive licensing. The growing aggressiveness of some technology-transfer offices in asserting their patents is now souring relationships between universities and industry, especially in information technology. Wayne Johnson, vice-president for university relations at computer giant Hewlett Packard in Palo Alto, California, testified before Congress in 2007 that Bayh–Dole has "fuelled mistrust, escalated frustration, and created a misplaced goal of revenue generation, which has moved universities and industry farther apart than they've ever been"[7].

Developing countries should not follow the United States in enacting policies that undermine traditional ways of commercializing research output. Patents and exclusive licences can boost technology transfer when significant follow-on investment is needed to promote commercialization — for instance, in the development of pharmaceutical compounds or prototypes for medical devices. But outputs that can be used off-the-shelf, such as computer software and biotechnology techniques, can be more effectively transferred by academic publishing, collaborations and teaching.

TECHNOLOGY TRANSFER?

University patents and licences have multiplied in recent decades, but this says little about the amount of technology transfer that is happening.



SOURCES: PATENT BOARD; NSF SRS; AUTM LICENSING SURVEY

In India, a version of Bayh–Dole-type legislation, drawn up in 2008, came close to mandating the patenting of all academic research output; institutions that did not comply would risk having their funding withdrawn. An outcry from academics and others has led policy-makers to soften their approach[8]. But the policy now under consideration still encourages patenting and licensing across the board — for example, for many of the software inventions emerging from Indian laboratories. In the Philippines, the recently passed Bayh–Dole analogue similarly fails to distinguish between inventions that should and shouldn't be patented, although regulations to control how the legislation is implemented are still being developed.

Indeed, policies promoting broad and aggressive patenting may be more of a problem in developing countries now than they were 30 years ago in the United States. More things are legally considered patentable, and under-resourced patent offices may struggle to weed out applications that aren't truly innovative. Legislators in developing countries need to distinguish between, and provide guidance on, the types of research that should be patented and exclusively licensed, and those that should be widely disseminated.

Countries considering Bayh–Dole type legislation should also be prepared to subsidize their technology-transfer offices. Few US universities are making large returns[9] and many make negligible income or even a net loss. One approach to this problem is for funders to allocate grant money for the management of intellectual-property rights for the types of research likely to need it.

A complicated issue for developing countries is whether they should treat academic patents and licences as a way to ensure that domestic firms and consumers, rather than large multinational companies, enjoy most of the benefits of taxpayer-funded research. This is particularly salient in countries without strong indigenous commercial capability, and where companies from developed nations might be better able to exploit innovations. Here again, drugs are a special case. For drug candidates with substantial markets in developed countries — those for 'global' diseases such as HIV or cancer — university licensing policies could be designed to simultaneously promote local access and preserve strong incentives for drug development[3].

There is no one-size-fits-all solution. Given the growing importance of developing-country universities in the global scientific enterprise, and the importance of public sector research for development, it is crucial that nations base their patent-reform laws on a clear-eyed assessment of their own problems and priorities. The choices made today will have profound effects on research, innovation and society for decades to come. ■

Bhaven N. Sampat *is in the Department of Health Policy and Management at Columbia University's Mailman School of Public Health, New York, New York 10032, USA.*
e-mail: bns3@columbia.edu

1. *Science and Engineering Indicators 2010* (NSF, 2010); available at go.nature.com/6tnlfx
2. Federal Council for Science and Technology *Annual Report on Government Patent Policy* (US Government Printing Office, 1976).
3. Sampat, B. N. *Am. J. Public Health* **99,** 9–17 (2009).
4. Cohen, W., Nelson, R. & Walsh, J. *Manage. Sci.* **48,** 1–32 (2002).
5. Rosenberg, N. & Nelson, R. *Res. Policy* **23,** 323–348 (1994).
6. Hughes, S. S. *Stanford's Office of Technology licensing and the Cohen/Boyer cloning patents* (The Bancroft Library, Univ. Calif. Berkeley, 1997); available at go.nature.com/xt2gj9
7. Johnson, W. *Bayh-Dole — the Next 25 Years.* Testimony to the Technology and Innovation Subcommittee of the Committee on Science and Technology (2007); available at go.nature.com/pgynki
8. Unnikrishnan, C. H. Parliamentary Panel Calls for Transparency in Public Research. *Mint* (10 August 2010); available at go.nature.com/tdwhus
9. *Managing University Intellectual Property in the Public Interest* (NRC, 2010).

P. J. RICHARDS/AFP/GETTY

More for the research dollar

Funders and universities should make the products of research more available — even if today's researchers pay a price, say **Jeffrey L. Furman**, **Fiona Murray** and **Scott Stern**.

On 29 October, the US government filed a brief stating that isolated but unmodified DNA should not be patented because merely isolating something does not turn it into a man-made product. This statement — submitted in a high-profile lawsuit over the validity of patents covering two genes linked to cancer — may or may not prevent the US Patent and Trademark Office adding new gene patents to the thousands already issued[1]. But the move has deepened the chasm between advocates of patenting research findings, and those calling for free and open access to publicly funded research.

The dispute over whether genes should be patented (or whether it is even legal to patent them) is typical of a wider debate. Researchers, open-source software designers, technology-transfer offices and entrepreneurs tend to fall into one of two camps with opposing opinions over whether patents, and intellectual-property rights over scientific findings, are 'good' or 'bad'.

Recent research in economics paints a far more complex picture. It suggests that scientific progress is not held up by intellectual-property rights per se but by the short-sighted ways in which these rights are often managed. It is time for scientists, universities and, in particular, funding agencies to start acting on such findings.

➲ NATURE.COM
US government to limit gene patents:
go.nature.com/nfqyos

The concept of 'governance' — the rules, expectations and practices through which people, organizations or resources are controlled — is as relevant to research institutions as it is to corporations. Thirty years ago this month, the Bayh–Dole Act altered the governance of science in the United States by replacing a confusing mass of rules over the ownership of patents with an overarching policy. The act gave universities — not funding agencies — the right to file and own intellectual property for inventions resulting from publicly funded research.

SLOWING PROGRESS

Economics is now beginning to shed light on the real-world effect of different governance schemes on scientific progress. Take a recent study[2] by Heidi Williams from the National Bureau of Economic Research in Cambridge, Massachusetts. This shows that, in the race to sequence the human genome, the different approaches to intellectual-property management adopted by Celera Genomics, then in Rockville, Maryland, and the Human Genome Project had a dramatic effect on the rate of follow-on research.

In the late 1990s, Celera Genomics, headed by Craig Venter, used copyright law to limit access to the firm's gene-sequence data. By contrast, the US-government-funded Human Genome Project made its data available with minimal restrictions. Using indicators such as patents, numbers of papers published and commercially available diagnostic tests, Williams compared the rate of research associated with genes sequenced by Celera to that associated with genes sequenced by the Human Genome Project. She found that a diagnostic test was 30% less likely to be developed for Celera-sequenced genes.

Another set of studies involving patents owned by Harvard University in Cambridge, Massachusetts, and the US chemicals company DuPont[3,4] demonstrates how shifts in governance can enhance scientific and technological progress. In the 1990s, DuPont required academics and researchers working for other companies to sign complex licensing agreements to use or develop two technologies used in mouse genetic engineering — the company's Cre-*lox* recombinant technology and the OncoMouse (a mouse strain modified to carry a cancer-causing gene developed at Harvard and exclusively licensed to DuPont).

Harold Varmus, then director of the US National Institutes of Health, established an agreement with DuPont in the late 1990s that changed how the company's patents were managed[5]. Clear, simple licensing guidelines, and low-cost access to the mice enhanced follow-on research and prompted a burst of activity in novel areas. Mice strains derived from these technologies were cited at a 30% higher rate over expected levels for several years after the policy change[4].

Other work suggests that restricting researchers' physical access to resources can be as damaging as doing so through contracts. For years, cell biology has been

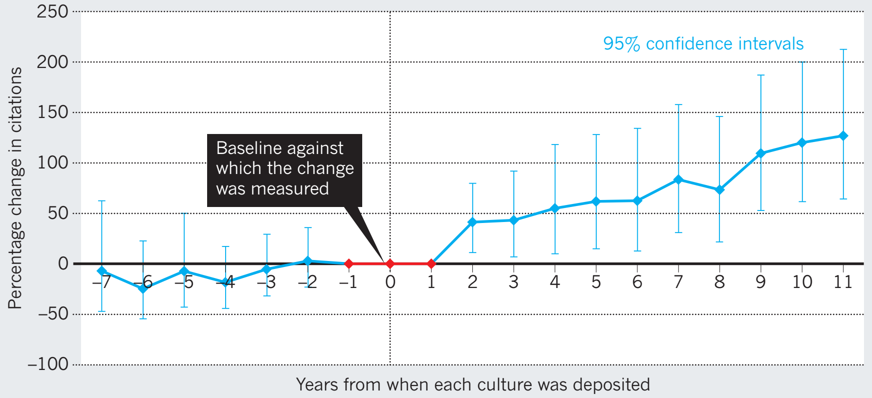
hampered by scientists storing cell lines independently. A recent study shows that the numbers of papers linked to 108 cell lines jumped more than 50% within 3 years of such lines being transferred to biological resource centres — such as the American Type Culture Collection in Manassas, Virginia[6] (see 'The positive effect of access'). Thus, even when biomaterials are unencumbered by intellectual-property rights, making them accessible through a trusted, open-access resource centre increases their effect on research.

The challenge is to provide incentives for today's researchers to create and characterize novel materials, models and databases, while ensuring that tomorrow's researchers can access and use these resources to enhance their own productivity. We recommend, first, that scientists and policy-makers establish rules of practice that maximize the productivity of research in the long term — even if those rules cost today's researchers some inconvenience or loss of competitive edge. The data-sharing strategy used by sequencers of the human genome offers a striking example of the effectiveness of this type of long-range planning.

In 1996, those involved in sequencing the human genome, including the US National Institutes of Health and the UK Medical Research Council, introduced the Bermuda Rules. These essentially require publicly funded researchers to deposit their sequencing data on a daily basis. Where researchers once had a monopoly over their data for several months, they now have sole access for less than 24 hours. In the short term, sequencers are less able to extract private value from their work. The benefits to subsequent research generations, however, in being able to quickly and easily access new sequence data soon after it is generated, have been enormous.

THE POSITIVE EFFECT OF ACCESS

The number of papers citing research on 108 cell lines rose rapidly after the cell lines were moved to a centralized, open-access culture collection. Data normalized by cell line, age of research and year of citation.



GREATER DISCLOSURE

Our second recommendation is that as a default, licensing transactions resulting from publicly funded research be disclosed. The results of research are generally made accessible through publishing, but materials — such as cell lines or tissue samples — and licensing contracts can be extraordinarily hard to obtain. For instance, at least nine patents owned by eight different entities[1] cover the *PSEN2* gene for a membrane protein. Although the information regarding ownership of intellectual-property rights is published by the US Patent and Trademark Office, neither universities nor companies publicize which companies have licensing contracts with the patent owners.

To change this, funding agencies should insist that licensors report each transaction including the identity of licensees and, when feasible, the structure of the transaction[7]. A standardized, accessible database of such transactions (managed perhaps by the US National Science Foundation) would reduce future transaction costs for innovators trying to build on ideas with many different patented elements.

Third, licences and other access rules should be structured so as to enable further research by as diverse a group of scientists, innovators and entrepreneurs as possible. This does not happen for many resources. For example, roughly 60% of university licences are awarded exclusively to single companies[8]. This means that scientists at other institutions or companies invariably have to pay for, or are prohibited from using, particular ideas[9]. It also means that it is up to licensees whether others can use the university's intellectual property to develop novel applications, or make resulting products available to the widest set of users, including in developing countries.

The Bayh–Dole Act grants universities flexibility in shaping how intellectual-property rights are used, but most funders are passive in ensuring that such rights don't inhibit cumulative research. In the OncoMouse case, for instance, policy-makers reacted only after a decade of dispute. Some technology-transfer offices have tried to come up with standard language for transparent licensing agreements to ensure, for example, global access to ideas and to the products generated from them. Although not yet widely adopted by universities, such an approach provides a valuable starting point.

Encouraging the broadest possible use of resources must apply to physical access as well. Some well-intentioned foundations, such as the International Myeloma Foundation (IMF) in North Hollywood, California, have taken the lead in establishing crucial disease-specific resources, including patient tissue samples. But, like the IMF, some foundations have granted only a select set of researchers access to the samples in the hope of attracting them to unique research opportunities. A better model is provided by the Coalition Against Major Diseases established by the Critical Path Institute, in Tuscon, Arizona. In June this year, the members — including patient advocates, pharmaceutical companies, and various institutes and agencies — agreed to pool and share data from failed Alzheimer's disease clinical trials, thereby broadening access to otherwise proprietary data.

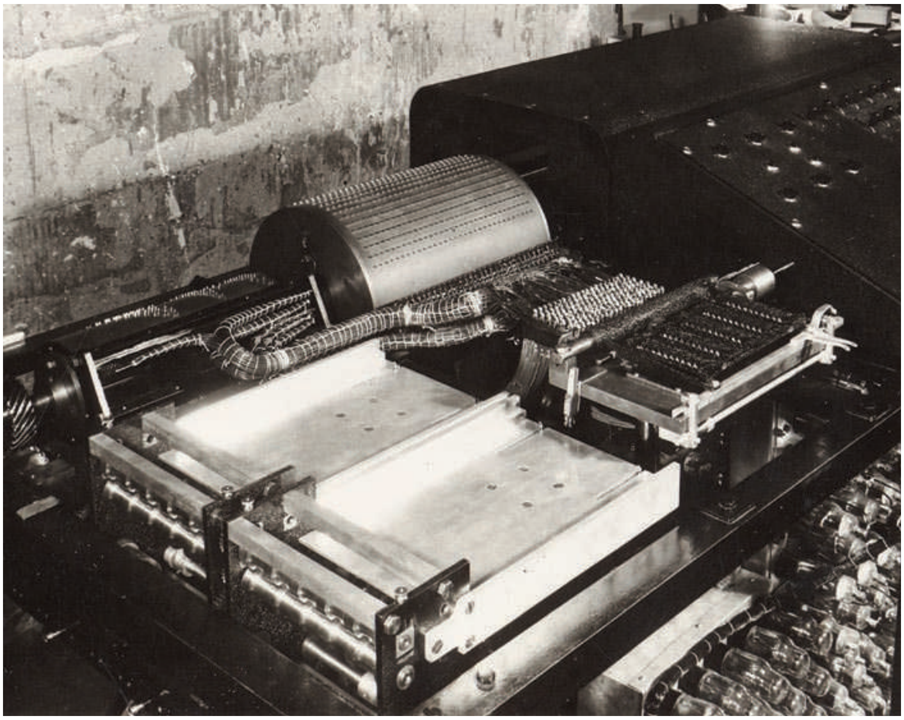
At a time when the public funding of science is under intense scrutiny, tremendous opportunity exists to establish policies that would greatly increase the impact of every dollar of research funding spent. ■

Jeffrey L. Furman *is at the Boston University School of Management, Boston, Massachusetts 02215, USA.* **Fiona Murray** *is at the MIT Sloan School of Management, Cambridge, Massachusetts 02142, USA.* **Scott Stern** *is at the MIT Sloan School of Management, Cambridge, Massachusetts 02142, USA.*
e-mail: fmurray@mit.edu

1. Jensen, K. & Murray, F. *Science* **310,** 239–240 (2005).
2. Williams, H. L. *Intellectual Property Rights and Innovation: Evidence from the Human Genome* NBER Working Paper 16213 (2010).
3. Murray, F. *Am. J. Sociol.* (in the press).
4. Murray, F., Aghion, P., Dewatripont, M., Kolev, J. & Stern S. *Of Mice and Academics: The Impact of Openness on Innovation* NBER Working Paper 14819 (2009).
5. Smaglik, P. *Nature* **403,** 350 (2000).
6. Furman, J. & Stern S. *Am. Econ. Rev.* (in the press).
7. Lemley, M. A. & Myhrvold, N. *Hofstra Law Review* **36,** 257–259 (2007).
8. Shane, S. & Somaya, D. *J. Econ. Behav. Org.* **63,** 739–755 (2007).
9. Chandrasekharan, S., Kumar, S., Valley, C. M. & Rai, A. *Nature Biotechnol.* **2,** 140–144 (2009).

CORRECTION

The Comment article 'Tar sands need solid science' (D. Schindler *Nature* **468,** 499–501; 2010) stated that the 650 km^2 footprint of the tar-sands mining is one-hundredth the size of Alberta or Texas. It is one-thousandth the size of those areas.



The 1940s Atanasoff–Berry Computer (ABC) was the first to use innovations such as vacuum tubes.

IOWA STATE UNIV. LIB./SPECIAL COLLECTIONS DEPT

BIOGRAPHY

The ABC of computing

An engaging biography of John Atanasoff reveals the obscure origins of the computer, explains **John Gilbey**.

Who invented the digital computer? Depending on your definition, mathematical pioneers such as John von Neumann or Alan Turing might spring to mind, but its origin lies with US physicist John Atanasoff. Although few people could name him today, this rewarding biography by Pulitzer prizewinning author Jane Smiley may change that.

Atanasoff embodies the American Dream. The son of a Bulgarian immigrant who had fled to the United States as a child in the late 1880s, he grew up on the family farm in Florida. Through mastering the slide rule, helping his father with house electrical wiring and driving the family's Model T Ford at age 11, he developed a passion for engineering and mathematics.

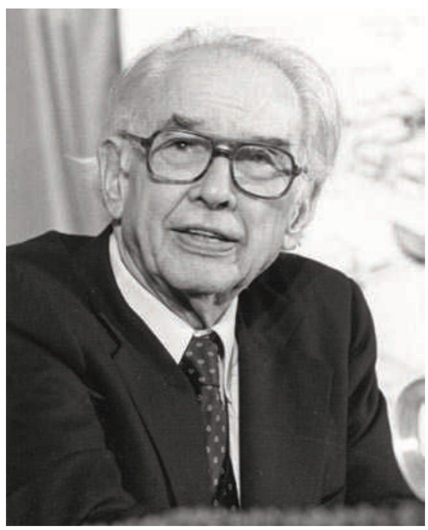
After graduating from the University of Florida in Gainesville in 1925, with the highest grade average it had ever recorded, Atanasoff joined a master's programme at what is now Iowa State University in Ames. He turned down an offer to move to Harvard University and gained a PhD in physics at the University of Wisconsin-Madison. He returned to Iowa State — again declining an offer from Harvard — as an assistant professor.

In *The Man Who Invented the Computer*, Smiley describes how Atanasoff developed an interest in mechanical calculators and modified an IBM tabulator to suit his own needs. But to meet his wider scientific aspirations — in particular, to solve simultaneous linear equations quickly — he realized that he would have to build a calculator himself. His struggle to design it concluded with an episode of pure cinema. Atanasoff, "unhappy to an extreme degree", jumped in his car and drove more than 300 kilometres to the shore of the Mississippi River. Sitting in a roadside tavern with a glass of bourbon and soda, the solution fell into place. He began to make notes on a paper napkin.

Crucially, Iowa State had an excellent college of engineering. In 1939, Atanasoff teamed up with recent graduate Clifford Berry to develop the system that became known as the Atanasoff–Berry Computer (ABC). Built on a shoestring budget, the simple 'breadboard' prototype that emerged contained significant innovations. These included the use of vacuum tubes as the computing mechanism and operating memory; binary and logical calculation; serial computation; and the use of capacitors as storage memory. By the summer of 1940, Smiley tells us, a second, more-developed prototype was running and Atanasoff and Berry had written a 35-page manuscript describing it.

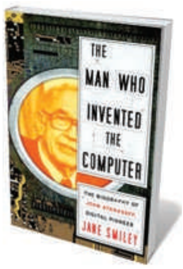
Other people were working on similar devices. In the United Kingdom and at Princeton University in New Jersey, Turing was investigating practical outlets for the concepts in his 1936 paper 'On Computable Numbers'. In London, British engineer Tommy Flowers was using vacuum tubes as electronic switches for telephone exchanges in the General Post Office. In Germany, Konrad Zuse was working on a floating-point calculator — albeit based on electromechanical technology — that would have a 64-word storage capacity by 1941. Smiley weaves these stories into the narrative effectively, giving a broad sense of the rich ecology of thought that burgeoned during this crucial period of technological and logical development.

The Second World War changed everything. Atanasoff left Iowa State to work in the Naval Ordnance Laboratory in Washington DC. His prototype computer remained unpatented in the basement of the physics department until the machine was broken up in 1948. The exigencies of war meant that substantial resources were made available for key computing projects such as the vast Electrical Numerical Integrator and Calculator (ENIAC) machine at the University of Pennsylvania in Philadelphia, the launch of which Atanasoff attended in 1946. But Atanasoff moved on, and in 1951 went into



AIP EMILIO SEGRÈ VISUAL ARCHIVES, *PHYSICS TODAY* COLLECTION

John Atanasoff built the first electronic computer.



The Man Who Invented the Computer: The Biography of John Atanasoff, Digital Pioneer
JANE SMILEY
Doubleday: 2010. 256 pp. $25.95

business for himself. His Ordnance Engineering Corporation was sold for a healthy profit five years later.

Atanasoff was brought back into the picture by the untimely death of Berry in an apparent suicide in 1963. Concerned, Atanasoff travelled to New York to investigate. The family considered that murder was a possibility — Berry's father had been shot decades earlier by a disgruntled ex-employee — but it was never proven.

In 1973, Atanasoff again found himself in the spotlight after his work was cited in the conclusions of a patent dispute between computing-industry giants Honeywell and Sperry Rand about the early development of the digital computer. Smiley quotes Judge Earl Larson's acknowledgement that "between 1937 and 1942, Atanasoff … developed and built an automatic electronic digital computer for solving large systems of simultaneous linear algebraic equations".

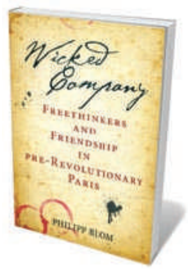
Judge Larson further noted that John Mauchly, one of the ENIAC developers who had visited Atanasoff in Iowa, had inspected the Atanasoff–Berry Computer and had read the manuscript describing it. Mauchly derived from this, the judge said, "'the invention of the automatic electronic digital computer' claimed in the ENIAC patent" — indicating Atanasoff's key contribution, albeit unwitting, to the later project.

Belatedly, and largely through the advocacy of friends and writers, Atanasoff gained recognition. Owing to his father's origins, he received early plaudits in Bulgaria, where in 1970 he was granted the Order of Cyril and Methodius, First Class. In 1990 he was awarded the National Medal of Technology by President George H. W. Bush for his invention of the electronic digital computer and for contributions to the development of a technically trained US workforce. Atanasoff died in 1995.

The Man Who Invented the Computer is a vivid telling of the early story of the computing industry. By focusing on Atanasoff, Smiley blends obscure threads with those that are better known. The result would, without embellishment, make an exceptional feature film. ■

John Gilbey *teaches in the Department of Computer Science at Aberystwyth University, Aberystwyth, Ceredigion SY23 2AX, UK.*
e-mail: gilbey@bcs.org.uk

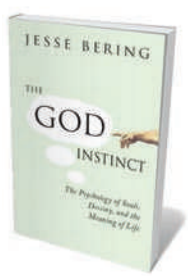
Books in brief



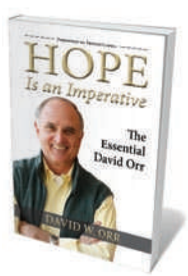
Wicked Company: Freethinkers and Friendship in Pre-Revolutionary Paris
Philipp Blom BASIC BOOKS *384 pp. $29.95 (2010)*
The French Enlightenment's triumph of reason over religious dogma was plotted in an eighteenth-century Paris salon. Hosted by Baron Paul-Henri Thiry Holbach, the radical thinkers who gathered there included the philosophers Denis Diderot and Jean-Jacques Rousseau. Historian Philipp Blom revives their legacy and examines the rivalries that sprang up among the group and with competitors such as the writer Voltaire. Their ideas about society and the natural world went on to influence politics and science globally.

How Old is the Universe?
David A. Weintraub PRINCETON UNIVERSITY PRESS *380 pp. $29.95 (2010)*
Astronomer David Weintraub explains in his latest book how we know that the Universe is 13.7 billion years old, a finding that has had an impact on science, philosophy and religion. By looking at the various ways in which the age of the cosmos has been established over the centuries — from the lifecycles and pulsations of stars to galactic structures and cosmology — he reveals the process of scientific enquiry and shows how astronomers gather evidence to grapple with deep questions.

The Abacus and the Cross: The Story of the Pope Who Brought the Light of Science to the Dark Ages
Nancy Marie Brown BASIC BOOKS *328 pp. $27.95 (2010)*
Far from being intolerant of science, the medieval Catholic Church saw reason as a means of getting closer to God. In the year 1000, there was even a 'scientist pope': Gerbert of Aurillac was the leading mathematician and astronomer of his day. Science writer Nancy Marie Brown describes his dramatic rise from humble peasant to visionary pontiff. A mathematics teacher to kings, and occasional spy, he adopted scientific ideas from the Islamic world, including the nine Arabic numerals and the concept of zero.



The God Instinct: The Psychology of Souls, Destiny and the Meaning of Life
Jesse Bering NICHOLAS BREALEY PUBLISHING *288 pp. £16.99 (2010)*
Psychologist Jesse Bering argues that religious beliefs are a sophisticated cognitive illusion rather than an irrational delusion. Because we have the ability to think beyond our immediate surroundings, we have evolved a tendency to project the idea that a transcendent being, or God, influences our lives. Taking a balanced and considered approach to this often inflammatory topic, he explains why this religious trait has evolutionary benefits and why it sets us apart from other animals.



Hope is an Imperative: The Essential David Orr
David Orr ISLAND PRESS *400 pp. £31 (2010)*
Key writings of environmental scientist David Orr from the past 30 years are collected in this volume. A champion of ecological design, Orr explains why it is important to educate people about sustainability, why university campuses should be green, and the environmental consequences of bringing children into the world. Leading a push within his own town of Oberlin, Ohio, to embrace green building practices, he reveals why he is both an optimist and a pragmatist.


ANNTHEA LEWIS

Sky Mirror (2006) in London's Kensington Gardens is one of four Anish Kapoor works manufactured by the process that is used to grind scientific optics.

SCULPTURE

Engineering art

Neil Dodgson admires the technical mastery of sculptor Anish Kapoor.

Indian-born British artist Anish Kapoor is famous for his architectural sculptures and vivid use of colour. His works are also feats of engineering — his speciality at university before he left to pursue his art. From *ArcelorMittal Orbit*, a tower of twisted steel chosen as the centrepiece for London's 2012 Olympic park, to *Svayambh*, a gliding slab of blood-red wax, the significance of Kapoor's installations lies in how they are built.

Kapoor, who is exhibiting in London, New Delhi and Mumbai, regards his sculptures as embodiments of "mythologies" that include the process of their creation. "Meaning is gradually constructed, just as the object is constructed," he explains. The shows in India highlight his dynamic artworks — shown in the past year at London's Royal Academy and at the Guggenheim Bilbao in Spain — which use machinery to evoke a sense of change. His current London exhibition, in Kensington Gardens, features four highly polished stainless-steel forms that distort reflections of their surroundings like fairground mirrors.

The genesis of that series lies in Kapoor's collaboration with Cecil Balmond, head of the Advanced Geometry Unit at engineering firm Arup. Kapoor first worked with Balmond a decade ago to produce a sculpture for the cavernous turbine hall at London's Tate Modern. He helped Kapoor to refine his aesthetic ideas, bringing expertise in construction techniques, the tensile strengths of materials and the limits of manufacturing. The product was *Marsyas* (2002) — two massive steel rings joined by a red PVC membrane stretched 140 metres between them, supporting a third steel toroid above visitors' heads. Balmond reprogrammed Arup's in-house software to model the membrane's precise form.

Anish Kapoor: Turning the World Upside Down
Kensington Gardens, London, until 13 March 2011.
Anish Kapoor
National Gallery of Modern Art, New Delhi, until 27 February 2011; and at Mehboob Studios, Mumbai, until 16 January 2011.

A discarded design later appeared in Chicago as *Cloud Gate* (2004): a 10-metre-high, jelly-bean-like arch of polished steel. The forms in Kensington Gardens are similar in style. Cut from segments of a sphere, they were produced by the same process that is used for grinding large scientific optics. One, *C-Curve* (2007), reminds me of a smaller mirror in my office: a relic of a prototype three-dimensional television. That too is beautifully made, but its bending of light is directed by a practical purpose. By contrast, Kapoor's curved mirrors are engineered to reflect the viewer's inner world.

In Kapoor's recent foray into dynamic works, now on show in India, exquisite engineering underlies other mythologies. In *Svayambh* (2007), an enormous block of red wax creeps along a hydraulic track, apparently being shaped as it passes through several gallery doorways. The name derives from a Sanskrit word, referring to that which is created of its own accord, rather than by a human hand. In fact, little wax is scraped off the installation after its first pass. Kapoor delights in this fiction: "The wax is not literally carved by the doorways, although it appears to be."

Kapoor's meanings are complex and layered. *Svayambh*, he explains, represents geology, body, blood and viscera, among other themes. It is difficult to engineer such a piece, with its combination of motors, mechanism and soft material requiring careful design and constant maintenance. Questioning the artist's intentions and methods unveils the fiction that the artwork formed itself.

A second wax piece seems more convincingly self-made. *Shooting into the Corner* (2008–09) is a large air-fired gun that fires 11-kilogram cylinders of red wax across the gallery every 20 minutes. The result is

a chaotic pile. No artist directs its creation; random perturbations are caused by variations in the consistency of the wax, the gun pressure and in how the deposits accumulate. Yet it is stage-managed. The art is not in the wax mound but in the whole performance.

Another artful machine generated a set of extruded grey concrete sculptures called *Greyman Cries, Shaman Dies, Billowing Smoke, Beauty Evoked* (2008–09). These were produced by a scaled-up version of a rapid prototyping machine. Such technology is normally used by engineers to build accurate models from fine threads of molten plastic. Kapoor's larger version extrudes a thick concrete sausage that builds up layers of soft coils, ropes and worms under computer control.

Whereas engineers seek precision with their models, Kapoor delights in his products' imperfections. Yet the appearance of randomness involves technical sleights of hand. To achieve each particular texture, his contraption must be finely tuned. Kapoor deliberately finds a point of balance between opposites — between perfection and imperfection, softness and firmness, movement and repose — to tantalize the viewer.

Kapoor knows he is treading a fine line between artist and entertainer. He says: "It's a short trip from Disneyland to something truly mysterious." But that mystery is delivered only through precise engineering. ■

Neil Dodgson *is professor of graphics and imaging at the University of Cambridge, UK.*
e-mail: neil.dodgson@cl.cam.ac.uk

BIBLIOMETRICS

Measure for measure

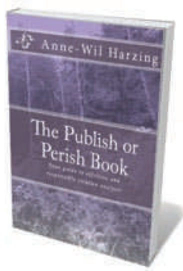
A useful guide to citation analysis shows that counting publications is harder than it looks, finds **Ton van Raan**.

Citation analysis offers a means to quantify the impact of a scientist's work. One tool for tracking citations is the Publish or Perish (PoP) software program developed by Anne-Wil Harzing, professor of international marketing at the University of Melbourne, Australia. Her guide describes how her program generates citation analyses from Google Scholar and gives an overview of bibliometric methods and sources. She champions the practical use of citation measures, yet also recognizes that calculating them reliably is a difficult task.

The Publish or Perish Book focuses on citation analysis of individual researchers, not groups or institutes. Several metrics may be calculated for scientists and for journals, including their number of publications and citations, average number of citations per publication and per author, and the h-index, a widely used characterization of citation impact. Harzing argues using practical examples that such indicators are good markers of a researchers' influence, and are useful in assessing applications for jobs, promotion and tenure, and for literature research and choosing a journal in which to publish.

Using Google Scholar as a data source is advantageous as it retrieves publications not covered by Thomson Reuters' Web of Science: books, edited volumes and 'grey' literature such as conference proceedings. Harzing explains how to analyse citations with Google Scholar, and discusses ways that citation patterns of early reports can be used to predict the later impact of journal articles derived from them. But there are inevitable problems in tying together varied data, such as matching conference proceedings with the subsequently published paper.

Harzing considers the main downside of the Web of Science to be its limited coverage of different disciplines, particularly of engineering, the social sciences and the humanities. In my view, however, its coverage of well-funded fields, such as the natural sciences and medicine, is very good. For novice users, the Web of Science does have trouble identifying ambiguous author names, especially those in which the order of the first name and surname is unclear. It also struggles to aggregate articles with variations of the same title and to identify self-citations. But professional bibliometricians such as myself build and work from Web of Science reconstructions — usually proprietary to their institutes — in which such sources of error are fixed.



The Publish or Perish Book: A Guide to Effective and Responsible Citation Analysis
ANNE-WIL HARZING
Tarma Software Research: 2010. 250 pp. $29.95

The book underplays the ethical issues that arise when performing a citation analysis for a person other than yourself. Verification of research output is important — missing just one highly cited paper can distort the results dramatically. This highlights the necessity of cleaning raw data. For instance, incorrect referencing will lead to cited publications being missed. It is a huge effort to correct for these 'homeless' citations. In this respect, Google Scholar is a black box.

Harzing discusses both the perspective of the person to be evaluated, and that of the evaluator. This is important because evaluators of tenure promotions might apply home-made metrics that are not transparent and may incorporate unknown mistakes. Citation metrics are attractive because they have the potential of objectivity, but evaluators may put too much faith in quantitative aspects of research performance. Simple metrics then become a shortcut to decision-making.

➲ NATURE.COM
For our Science Metrics special, see:
go.nature.com/nj2xqk

Many scientists are concerned that citation analysis, particularly that done in an amateurish way, is having detrimental effects on science. They fear that researchers are driven to pursue citation quantity instead of scientific quality. Statistical reliability may become a serious problem when dealing with individuals rather than groups, as Harzing recognizes. Field-specific normalization is also necessary if research impact is to be compared across disciplines.

Further statistical factors limit metrics. Indicators often concern arithmetic mean values, yet the distribution of citations across publications is skewed. Averages are thus not the best statistic. Although this problem is discussed, Harzing's book doesn't offer indicators that are related to the distribution of impact across a field, which would answer the question 'Does he or she belong to the top 10% of his or her field?'

Harzing explains that the problem of the skewed distribution can be removed using the h-index: for instance, a researcher has an h-index of ten if ten of his or her papers have at least ten citations and the other papers have no more than ten citations each. But in my view, the h-index is inconsistent. For example, suppose that researcher A has three publications with five citations each (h=3) and researcher B has four with four citations each (h=4). Both obtain one additional publication with five citations. Researcher A's h-index then increases to four, whereas researcher B's h-index remains equal to four. This makes no sense.

With these caveats in mind, *The Publish or Perish Book* is a useful resource for scientists, particularly in fields in which Google Scholar is a major source of citations. ■

Ton van Raan *is professor of science studies in the Centre for Science and Technology at Leiden University, the Netherlands.*

CONSERVATION

A magical process

Prince Charles's call to stay close to nature follows a rich tradition of environmental thinking, says **Philip Stott**.

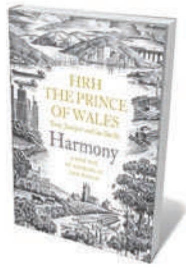
This esteemed journal notwithstanding, one should generally be wary of those who present Nature with a capital N. Such commentators resent the indifference they perceive in the natural world to moral values, preferring to see in nature a reflection of their own persona. They reify and then deify the natural world, and worship it as virtuous.

This is absolutely the tenor of *Harmony*, Prince Charles's call to virtue through humans emulating "the natural order and the rhythm in things". Like others before him, the heir to the British crown — who has long expressed views on the environment, and is a champion of organic and traditional farming methods — wishes to speak on nature's behalf. By taking as our guide the rhythms and patterns that lie within us, he writes, we may build a more durable and pleasant society and acquire deep philosophical insights that are embedded in our traditions.

For him, Nature's virtue lies in its ability to replenish itself efficiently and without waste, through cycles that amount to a "magical process". This is a fascinating concept, until one remembers that more than 95 per cent of all life has been discarded on an Earth where volcanoes blow, robins kill robins, forests come and go, and viruses prey on all.

Harmony clearly represents a personal statement by the prince. Although he thanks his co-authors, environmentalist Tony Juniper and radio broadcaster Ian Skelly, many paragraphs open with a resounding "I". Charles's mission is to articulate his belief that our broken connection with nature will drive humanity to oblivion. The cover states that "our disconnection from Nature has contributed to the greatest crisis in the history of mankind".

Such a jeremiad is far from new. The prince is rehearsing environmental themes that have a deep pedigree in European and American thought. The influential Vermont diplomat George Perkins Marsh, an early prophet of environmental concerns who wrote the 1864 masterpiece *Man and Nature*, would immediately recognize its tropes. So too would twentieth-century environmental writers such as Aldo Leopold, with his land ethic in *A Sand County Almanac* (1949), and H. J. Massingham with *The Wisdom of the Fields* (1945) — both published in the decade in which the prince was born. Charles would no doubt approve of Massingham's call for a "design for living" and support his conclusion that our "failure is so pronounced that it is dragging Western civilisation nearer and nearer to some fall like Lucifer's".



Harmony: A New Way of Looking at Our World
HRH THE PRINCE OF WALES WITH TONY JUNIPER AND IAN SKELLY
HarperCollins: 2010. 336 pp. £25, $29.99

All of these tracts rest on long-standing European foundations. These include the idea of the 'virtuous rural' expressed in the Roman poet Virgil's *Eclogues* and *Georgics*, and the concept of the noble savage expounded in the late sixteenth and seventeenth centuries by French essayist Michel de Montaigne and English poet John Dryden. They also have a basis in German romanticism, including the philosophies of 'the organic' and holism developed in the nineteenth century by, for example, biologist Ernst Haeckel. As one might observe, there is little new under the solar panel.

The benevolent depiction of Nature so admired by the prince is conjured in his own well-meaning image. The cover of the UK edition — a charming set of neopastoralist cameos reminiscent of the inside covers of a classic Rupert Bear Annual — says it all. But for many, Nature is no place to seek an explanation of ethics, virtue or a sense of the numinous. There is little harmony in the modern ecological concepts of disequilibrium and non-equilibrium, which hold imbalance and constant change to be the essential state, nor in a different selection of philosophies and poems. Tennyson's powerful 1849 poem *In Memoriam*, for example, tells of Nature's inherently destructive character: "I care for nothing, all shall go".

The prince muses on issues of long-term concern to him — the threat of global warming, the promise of alternative medicines, the dangers in modern farming, the brutalities of urbanism and, like Massingham, the wisdom of the fields and of the past. There are inconsistencies. For example, the classic doom-laden picture of a tropical forest seemingly laid waste by 'slash and burn' agriculturalists inevitably appears. Yet such interventions can make use of the traditional approaches favoured by the prince — many practitioners of shifting cultivation use field cycles that preserve soils on steep slopes and increase yields. By contrast, some First Nation peoples were highly destructive of their environments, not protective.

Harmony is a mishmash of selected concepts of a reified Nature, rather a lot of mysticism and, admittedly, a selection of sound and welcome practical comments on aspects of farming and urban living. The prince is right to castigate our pollution of the oceans with plastic rubbish. It is surely a sick bird that fouls its own nest. Many will also welcome his support for the conservation of the red squirrel in the United Kingdom.

This attractively produced book may delight and stir those for whom the world has never been modern. For others, it might merit Viscount Castlereagh's acerbic dismissal of Tsar Alexander I's Holy Alliance of 1815: a "piece of sublime mysticism and nonsense". ■

Prince Charles advocates a return to the natural order of things.

TIM GRAHAM PICTURE LIBRARY/GETTY

Philip Stott *is emeritus professor of biogeography in the University of London, UK.*
e-mail: sinfonia1@mac.com

CORRESPONDENCE

Ease public concern over oil pipeline

The need for environmental scientists to address the effects of tar-sands mining in Canada (*Nature* **468,** 476; 2010) should be extended to the impacts of downstream operations.

Communities across the northern High Plains region of the United States are concerned about the risks of piping crude oil from Canadian tar sands across ecologically sensitive prairie and through an important recharge zone of the Ogallala Aquifer — the route of the proposed Keystone XL pipeline through the Nebraska Sand Hills.

The public debate is being conducted largely in the absence of scientific evidence about risks to water resources and aquatic ecosystems. This causes misinformation to circulate: for example, local stakeholders commonly believe that any oil released from a ruptured pipeline could contaminate the entire High Plains groundwater supply — based on the widespread misconception of an aquifer as an underground lake. Others, by contrast, believe that spilled oil would be harmlessly sequestered in the aquifer.

Much of the blame for these misconceptions must be down to poor communication with the public by scientists. However, scientists themselves are often hampered from providing technical input because of their limited access to data — as has happened with the Keystone XL proposal. Important data pertaining to this have not been divulged to the public, such as the fluid chemical composition and the maximum pipeline leakage volumes.

Disclosure of relevant data must be comprehensive if the risks associated with the pipeline are to be properly assessed.

John B. Gates *University of Nebraska-Lincoln, USA.*
jgates2@unl.edu

Sustainable cities: seeing past the trees

The problems facing our cities call for a holistic approach, not just for ecological solutions (*Nature* **468,** 173; 2010). We also need to consider the resilience of the changes we make to the urban landscape in the name of sustainability (see www.urban-futures.org) and strike a balance between the benefits and disadvantages of these strategies.

Take street trees planted to improve biodiversity. They reduce air pollution by increasing particle deposition and replenishing oxygen, yet may also exacerbate it by reducing ventilation. They provide shade but limit passive solar heating. Their amenity value may be undermined by high costs for repairing infrastructure damaged by ground shrinking and swelling. Although they help to mitigate light pollution, trees are likely to increase lighting requirements, and although they store water they may need irrigating — and so on.

A continuing positive outcome will depend on thoughtful assessment of the competing and shifting aspects of sustainability.

Rob MacKenzie, Tom Pugh *Lancaster University, UK.*
r.mackenzie@lancs.ac.uk
Chris Rogers *University of Birmingham, UK.*

Time to underpin Wikipedia wisdom

Wikipedia, the world's largest online encyclopaedia, is regarded with suspicion by some in the scientific community — perhaps because the wiki model is inconsistent with traditional academic scholarship (*Nature* **468,** 359–360; 2010). But the time has come for scientists to engage more actively with Wikipedia.

Type any scientific term into any search engine and it is likely that a Wikipedia article will be the first hit. Ten years ago, it would have been inconceivable that a free collaborative website, written and maintained by volunteers, would dominate the global provision of knowledge. But Wikipedia is now the first port of call for people seeking information on subjects that include scientific topics. Like it or not, other scientists and the public are using it to get an overview of your specialist area.

Wikipedia's user-friendly global reach offers an unprecedented opportunity for public engagement with science. Scientists who receive public or charitable funding should therefore seize the opportunity to make sure that Wikipedia articles are understandable, scientifically accurate, well sourced and up-to-date.

Many in the scientific community will admit to using Wikipedia occasionally, yet few have contributed content. For society's sake, scientists must overcome their reluctance to embrace this resource.

Alex Bateman, Darren W. Logan *Wellcome Trust Sanger Institute, Hinxton, UK.*
agb@sanger.ac.uk

Guest authors: for contributors only

In your Careers feature on tenured academic positions (*Nature* **468,** 123–125; 2010), you recommend that those seeking tenure should "Name a senior department member as a co-author on your papers if you're in Europe". But this regretfully common practice should be weeded out, not encouraged.

Such a recommendation would be worthwhile if it were to motivate a genuine collaboration and result in a significant contribution from the senior scientist to the paper.

Mark J. van Raaij *Centro Nacional de Biotecnologia, Spain.*
mjvanraaij@cnb.csic.es

Guest authors: no place in any journal

In the pursuit of tenure, you encourage academics in Europe to include a senior department member as co-author on their papers (*Nature* **468,** 123–125; 2010). Regardless of any geographical limitation, it was very disappointing to see this endorsement of guest authorship published in *Nature*.

Richard M. Glass Journal of the American Medical Association, *Chicago, Illinois, USA.*
richard.glass@jama-archives.org

Editor's note *Nature* requests author contributions statements and does not condone guest authorship (see go.nature.com/uzgbki). The intended meaning of the ambiguous sentence was that tenure-track scientists should seek out senior department members as collaborators and contributing co-authors.

CONTRIBUTIONS

Items for Correspondence may be sent to **correspondence@nature.com** after consulting the author guidelines at **http://go.nature.com/cmchno**. They should be no longer than 350 words. Readers are also welcome to comment online on anything published in *Nature*: **www.nature.com/nature**.

Michael Tinkham

(1928–2010)

Physicist who helped to unravel the mysteries of superconductivity.

Almost 100 years after superconductivity was discovered in 1911, the field has lost one of its finest contributors and certainly its most important contemporary articulator of how the phenomenon works.

Throughout his life, Michael Tinkham, who died on 4 November, never lost his remarkable ability to recognize the essentials and explain them to the rest of us. A colleague of his once said you could take him data that looked like pigeon droppings and leave with flakes of gold. Irreverent perhaps, but legions of graduate students and postdocs shared this experience. Tinkham's classic book *Introduction to Superconductivity*, first published in 1975, remains to this day the definitive treatment of the subject— making it accessible to a wide range of scientists and engineers.

Tinkham started his academic career as an undergraduate at Ripon College in Wisconsin, near where he grew up. After graduating, he went to the Massachusetts Institute of Technology in Cambridge, where he received his master's and his PhD, before completing a postdoc at the University of Oxford, UK.

When he returned to the United States in 1955, he took a faculty position at the University of California, Berkeley. It was here that he began to develop his life-long interest in superconductivity — the astonishing property that some metals have at very low temperatures of allowing current to flow through them with no resistance.

"You could take him data that looked like pigeon droppings and leave with flakes of gold."

Tinkham recognized that many properties of solids might usefully be studied spectroscopically using far-infrared radiation — until this point, its use had been confounded by poor sources and detectors. Specifically, he suspected that changes in the absorption of this radiation would occur when solids become magnetic or superconductive.

In the same way that electrons in atoms have energy levels, so do conventional metals. However, for metals, the distribution of these energy levels is essentially continuous. A characteristic property of a superconductor is that an energy gap forms in this continuous distribution. We now know that this gap results from electrons in the metal binding into pairs — a phenomenon fundamental to superconductivity — but in the mid-1950s, there wasn't even direct evidence that such a gap existed.

S. ROSNER

In 1956, Tinkham and fellow postdoc Rolfe Glover found the first direct evidence for this energy gap in the form of a sharp rise in the absorption spectrum of a superconductor. They also noted that aspects of the absorption data were counter-intuitive. For instance, the amount of radiation absorbed didn't just steadily rise as Tinkham and his group increased the energy of the radiation beyond that needed for any absorption to occur. It rose above the level of absorption one might expect for a non-superconducting metal before decreasing to the expected value. Tinkham loved to tell the story of how, when he mentioned these peculiar observations to John Bardeen, who was already working with Leon Cooper and Robert Schrieffer on what turned out to be the correct theory of superconductivity (the BCS theory, proposed in 1957), Bardeen simply commented that such behaviour was "not unexpected".

Bardeen was right: these observations were a direct consequence of the celebrated 'coherence factors' of the BCS theory. Bardeen and his colleagues soon established that when superconducting electron pairs are broken apart (for example, by radiation), instead of producing two separate electrons, a combination of electrons and 'holes' results — a 'hole' being the conceptual and mathematical absence of an electron. This observation, along with other unusual phenomena, measured for instance by passing sound waves through superconductors, provided the first substantive experimental confirmation of the BCS theory. Bardeen, Cooper and Schrieffer went on to win the 1972 Nobel Prize in Physics.

REAL-WORLD EFFECTS

This spectacular role in the early history of the BCS theory behind him, Tinkham continued to work with far-infrared spectroscopy but also began to study the macroscopic quantum behaviour of superconductors. Quantum mechanics is normally thought of as important only in the microscopic world of atoms, but in superconductors it manifests itself in very large objects, such as in the superconducting magnets used in magnetic resonance imaging. After Tinkham took up a professorship in 1966 at Harvard University in Cambridge, Massachusetts, one question emerged that remained of central interest to him: what is the nature and origin of resistance in a superconductor? Or put more simply, when is a superconductor really a superconductor?

As it turns out, when superconductors are carrying a current, they don't stay in a fixed macroscopic quantum state, but cascade down from one energy level to another. As energy is lost with each transition, this is equivalent to saying that superconductors have resistance, although it is extremely small under most conditions. In the latter stages of his career, Tinkham was examining the conditions under which these transitions happen and how they happen in very thin wires of a superconductor.

Despite all these achievements, being elected to the US National Academy of Sciences and winning the prestigious Oliver E. Buckley prize of the American Physical Society, Mike was a modest man with an exceptional sense of humour. The same legions of students who witnessed his alchemy with data will remember how they first knocked nervously on his door to be greeted by a somewhat gruff "Come in", only to learn that he was a very warm and witty mentor.

I was privileged to be in Mike's group at Harvard from 1968 to 1974. His breadth of skills — as part theorist, part experimentalist — made him seem to me the archetype of a complete physicist. ■

Malcolm Beasley *is an emeritus professor of applied physics at Stanford University, Stanford, California 94305, USA.
e-mail: beasley@stanford.edu*

NEWS & VIEWS

EVOLUTIONARY BIOLOGY

Genomic hourglass

Comparative genomics studies reveal molecular signatures of the controversial 'phylotypic' stage — a time when embryos of members of an animal phylum all look more alike than at other embryonic stages. SEE LETTERS P.811 & P.815

BENJAMIN PRUD'HOMME & NICOLAS GOMPEL

Most people would say that lizards and elephants bear little resemblance to each other. But not so the embryologist, for, at a particular stage in development, the embryos of very different species may look much the same. Elsewhere in this issue, papers by Kalinka *et al.*[1] and Domazet-Lošo and Tautz[2] offer a fresh perspective on this intriguing phenomenon.

This is a topic with a long history. In 1828, the German biologist Karl von Baer, one of the fathers of embryology, reported how very similar the early embryos of different species can be[3]: "I have two small embryos preserved in alcohol, that I forgot to label. At present I am unable to determine the genus to which they belong. They may be lizards, small birds, or even mammals." In fact, it was later observed that, over the course of development, the youngest embryos within an animal phylum often look very different, but progressively converge towards a similar form (described by von Baer and later dubbed the phylotypic stage), before they diverge again to achieve the tremendous diversity of adult forms.

This pattern of morphological divergence among species during embryonic development resembles an hourglass[4,5]. Its waist marks the phylotypic period during which the basic body plan of a given animal group is laid down. The existence and meaning of the hourglass model, however, have been the subject of heated controversy, in part because the model rests on subjective comparisons of animal likeness of shape[6–9]. The contribution of Kalinka *et al.*[1] and Domazet-Lošo and Tautz[2] is to report molecular signatures supporting the existence of the phylotypic stage in insects and vertebrates.

To test the hourglass model, Kalinka *et al.*[1] (page 811) reasoned that, because the development of shape is directed by the expression of genes, variations in morphological pattern among species might be reflected in the dynamic of gene expressions. The authors set out to test this idea by measuring differences in gene expression between various species of the fruitfly *Drosophila*. Using DNA microarrays, a technology that measures genome-wide gene expression, they first quantified levels of gene expression throughout embryogenesis for six distinct species of *Drosophila*. Next, they compared the temporal expression profiles of all the genes across the six species.

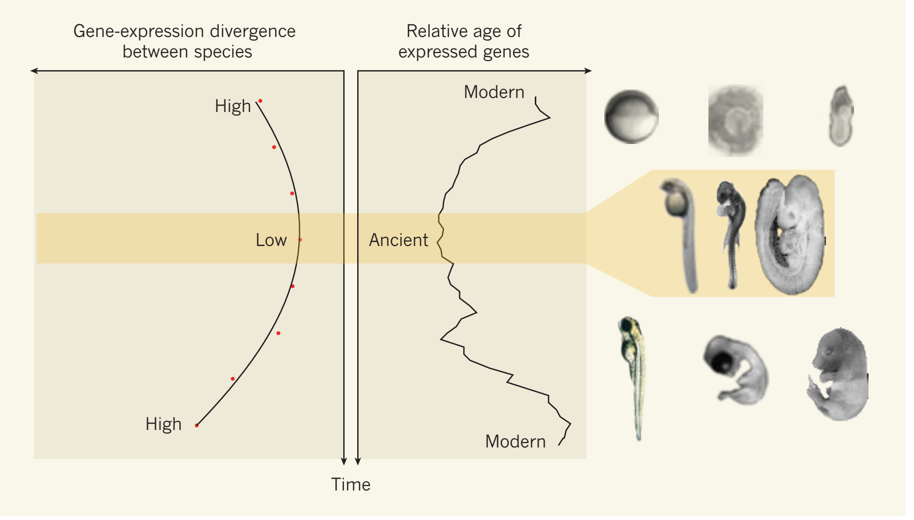


Figure 1 | The developmental hourglass, as revealed by comparative genomics. Mid-embryogenesis is marked by the phylotypic stage, a period of minimal anatomical divergence between species, as illustrated for vertebrate species by the orange band. This stage is now shown by Kalinka *et al.*[1] to display minimal gene-expression divergence between *Drosophila* species (left curve), and by Domazet-Lošo and Tautz[2] to express the oldest gene set of the entire life cycle (right curve). The species depicted, left to right, are zebrafish, chick and mouse. (Images reproduced from refs 12–14.)

A sophisticated statistical analysis of the data set revealed a pattern strikingly similar to the anatomical hourglass, in which the temporal gene-expression divergence among species is minimal around the 'extended germband' stage, which is classically regarded as the phylotypic stage in insects[10] (Fig. 1). That is, the expression of genes that are active during the extended germband stage is evolutionarily more stable than that of genes active earlier and later during development. Remarkably, the genes that mostly conform to an hourglass pattern are those involved in developmental processes, whereas genes involved in non-developmental functions show more variable expression profiles across species.

The detection of a phylotypic stage at the gene-expression level revives some long-standing considerations on the relationships between embryonic development (ontogeny) and evolution (phylogeny)[11]. Ontogeny clearly does not recapitulate phylogeny, yet these two processes have intricate connections. It is precisely the nature of these connections that Domazet-Lošo and Tautz[2] (page 815) have explored.

Starting from the notion that developmental novelties might be enabled by the evolution of new genes, these authors sought to correlate the emergence of new genes (or gene families) with novelties in the anatomical development of a species — zebrafish, in their case. They used 'phylostratigraphy', an approach they had developed previously, to parse the genome into classes of genes according to their evolutionary origin in the history of life ('phylostrata'). For instance, the zebrafish genome includes genes that date back to the origin of cells, others that date to the evolution of animals and yet others that date to the evolution of vertebrates. Then, using DNA microarrays, the authors measured the relative contribution of each phylostratum to global gene expression (the 'transcriptome') at different time points in the zebrafish life cycle, thereby estimating the relative age of the transcriptome at each time point.

It turns out that genes of different evolutionary origins are expressed at different time

points (Fig. 1). Strikingly, the stage classically viewed as the phylotypic stage in zebrafish is marked by the expression of the evolutionarily oldest transcriptome set, whereas earlier and later stages (including adult stages) express comparatively younger transcriptomes. Importantly, the authors identified a similar pattern in published microarray data for other organisms (fruitfly, mosquito and nematode), suggesting that their findings are generally applicable.

By revisiting the subjective anatomical comparisons of classical embryology using quantitative genomics, these two studies[1,2] have revived the concept of the phylotypic stage with much-needed objectivity. Although they take very different approaches, it is remarkable that both studies identify genomics signatures of the phylotypic stage — in short, the phylotypic stage sees expression of the oldest gene set, which is maximally conserved across species. These results reinforce the notion that animal body plans emerged using novel signalling and regulatory genes that arose at the inception of multicellular animal life, and that, once established, the gene-expression patterns underlying the specification of the different body plans have remained fairly invariant.

This newly acquired molecular legitimacy does not, however, explain what establishes and maintains the hourglass pattern. Kalinka *et al.*[1] found that the hourglass pattern of gene-expression variation is best explained by the action of natural selection. This echoes the proposition that mechanistic constraints pertaining to the building of a shared body plan might explain the conservation observed at the phylotypic stage[4,5].

A body plan is a particular organization of anatomical rudiments. The early embryonic specification of these rudiments, independently of one another, might take different evolutionary roads. But the assembly of these elements into a functional body plan might require a tight and constrained orchestration of gene expression, reflected in the hourglass waist. Once coherently assembled, the connected elements make a stable evolutionary substrate for an organism to explore new morphogenetic directions within the realm of the established body plan.

With this work[1,2], new avenues open up in addressing a long-standing debate. Future comparative studies of the gene-regulatory networks and developmental events underlying the phylotypic stage will certainly shed light on the *raison d'être* of this peculiar embryonic period. ■

Benjamin Prud'homme *and* **Nicolas Gompel** *are at the Institut de Biologie du Développement de Marseille-Luminy, Parc scientifique de Luminy, case 907, 13288 Marseille cedex 9, France.*
e-mails: benjamin.prudhomme@univmed.fr; nicolas.gompel@univmed.fr

1. Kalinka, A. T. *et al. Nature* **468,** 811–814 (2010).
2. Domazet-Lošo, T. & Tautz, D. *Nature* **468,** 815–818 (2010).
3. Scott, G. F. *Developmental Biology* 6th edn, 9 (Sinauer, 2000).
4. Duboule, D. *Development* (Suppl.) 135–142 (1994).
5. Raff, R. A. *The Shape of Life: Genes, Development and the Evolution of Animal Form* (Univ. Chicago Press, 1996).
6. Bininda-Emonds, O. R. P., Jeffery, J. E. & Richardson, M. K. *Proc. Biol. Sci.* **270,** 341–346 (2003).
7. Hall, B. K. *Trends Ecol. Evol.* **12,** 461–463 (1997).
8. Hazkani-Covo, E., Wool, D. & Graur, D. *J. Exp. Zool. B* **304,** 150–158 (2005).
9. Richardson, M. K. *et al. Anat. Embryol. (Berl.)* **196,** 91–106 (1997).
10. Sander, K. in *Development and Evolution* (eds Goodwin, B. C., Holder, N. & Wylie, C. C.) 137–159 (Cambridge Univ. Press, 1983).
11. Gould, S. J. *Ontogeny and Phylogeny* (Harvard Univ. Press, 1977).
12. Kimmel, C. B. *et al. Devl Dynam.* **203,** 253–310 (1995).
13. Hamburger, V. & Hamilton, H. L. *Devl Dynam.* **195,** 231–272 (1992); reprinted from *J. Morphol.* **88,** No. 1 (1951).
14. Mouse Atlas Project, MRC Human Genetics Unit www.emouseatlas.org/Atlas

QUANTUM PHYSICS

Hot entanglement

Quantum entanglement has been observed at low temperatures in both microscopic and macroscopic systems. It now seems that the effect can also occur at high temperatures if the systems are not in thermal equilibrium.

VLATKO VEDRAL

Quantum physics is usually thought to apply to small systems at low temperatures. A standard example would be the quantum dynamics of an electron in a hydrogen atom. Atomic orbits of electrons are roughly an ångström in size — that is, comparable with electronic de Broglie wavelengths, which characterize the extent over which electrons display a quantum wave-like behaviour. More importantly, at low temperatures, the typical energies characterizing electronic jumps are hundreds of times larger than the thermal energy of the environment to which the system is exposed. This, in turn, means that the noise due to the environmental temperature is negligible compared with the typical electronic-jump energies, and therefore that the noise does not spoil the system's quantum behaviour. Writing in *Physical Review Letters*, Galve *et al.*[1] show that, contrary to the common view, a macroscopic system at high temperatures can also sustain quantum features.

It is interesting that similar considerations about the restriction of quantum phenomena to small systems at low temperatures can be made about the most quantum of all quantum effects: quantum entanglement. The term entanglement was coined by Erwin Schrödinger, who described it as "the characteristic trait of quantum mechanics". It refers to a state of two or more quantum systems in which the systems are so intertwined that they behave like one — it is actually a mistake to think of the subsystems separately. Quantum systems become entangled when they interact with one another. In the past decade, extensive theoretical and experimental research[2] has shown that, no matter what systems we look at, a general rule says that if the interaction strength between the subsystems is larger than the thermal energy due to their coupling to the environment, entanglement should exist between these subsystems provided that they are in thermal equilibrium with the environment.

Now Galve *et al.*[1] prove that this relationship between temperature and entanglement is not valid for systems that are not in thermal equilibrium. Here, in fact, the news is very good for entanglement. The authors predict that nanomechanical oscillators can be entangled at much higher temperatures than previously thought possible.

The basic intuition behind this result is as follows. When a system is not in thermal equilibrium, the temperature no longer provides the relevant energy scale against which to compare the system's quantum behaviour. What matters instead is an effective temperature, which can be much lower than the absolute one. This effective temperature is obtained by multiplying the absolute temperature by the rate at which the system approaches equilibrium divided by the driving frequency, the frequency of the signal with which the system is made to oscillate. Galve and colleagues demonstrate that this new condition for entanglement — that the interaction between subsystems should be compared with the thermal energy at the effective temperature — holds quite generally and is intuitively pleasing. It says that if we can drive the system to oscillate within a shorter timescale than the time it takes to reach thermal equilibrium, then an entangled steady state can be attained at higher temperatures than the absolute one.

The actual system that Galve *et al.* investigate — two macroscopic (harmonic) oscillators coupled to each other — is important because a number of laboratories are currently working with similar systems. For instance, Aspelmeyer and colleagues[3] have created quantum states

in a movable nanomechanical mirror that is a microgram in weight. The high-temperature entanglement envisaged by Galve *et al.* could be achieved by coupling two such mirrors to one another. I and colleagues[4] have shown, using a different theoretical approach to that of the present study[1], that such nanomechanical entanglement should persist at temperatures of about 20 kelvin. The hope now is that, by using Galve and colleagues' new ideas, the temperature can be pushed upwards to, say, 100 kelvin. This would eliminate the current need for expensive and elaborate cryogenics to cool the oscillators.

So, OK, we can in principle entangle nanomechanical oscillators at high temperatures. Physicists will no doubt get excited because this realization will strengthen the evidence for the universality of quantum mechanics. But why should anybody else care?

The most exciting macroscopic and 'hot' non-equilibrium systems we know are, of course, the living ones. We can, in fact, view any living system as a Maxwell's demon, maintaining life by keeping its entropy low against the environmental noise — that is, by being far from equilibrium. The father of thermodynamics, Ludwig Boltzmann, himself viewed living systems in this way. Here is what he said on the matter: "The general struggle for existence of living beings is therefore not a fight for energy, which is plentiful in the form of heat, unfortunately untransformable, in every body. Rather, it is a struggle for entropy that becomes available through the flow of energy from the hot Sun to the cold Earth. To make the fullest use of this energy, the plants spread out the immeasurable areas of their leaves and harness the Sun's energy by a process as yet unexplored, before it sinks down to the temperature level of our Earth, to drive chemical syntheses of which one has no inkling as yet in our laboratories."

We have actually learnt a little bit about that "unexplored" process — photosynthesis — since Boltzmann. And as it happens, recent experiments[5] show a quantum effect leading to entanglement[6] in some photosynthetic complexes. Such entanglement might yield an increased efficiency in the transfer and processing of energy in photosynthesis. The overall mystery of photosynthesis remains, but there is now evidence that quantum physics has something to do with it in a profound way. And there are other instances in biology in which quantum entanglement could be important[7]. If this is a general trend in the biological world (and it is a big 'if'), maybe Boltzmann was only half right: could it be that life does not just keep its entropy low, but rather, also aims to keep its quantum entanglement high if and when needed for an increased efficiency of energy transport? For now, the jury is still out. ■

Vlatko Vedral *is at the Clarendon Laboratory, University of Oxford, Oxford OX1 3PU, UK, and the Centre for Quantum Technologies, National University of Singapore, 117543 Singapore.*
e-mail: vlatko.vedral@qubit.org

1. Galve, F. *et al. Phys. Rev. Lett.* **105,** 180501 (2010).
2. Amico, L., Fazio, R., Osterloh, A. & Vedral, V. *Rev. Mod. Phys.* **80,** 517–576 (2008).
3. Gröblacher, S., Hammerer, K., Vanner, M. R. & Aspelmeyer, M. *Nature* **460,** 724–727 (2009).
4. Vitali, D. *et al. Phys. Rev. Lett.* **98,** 030405 (2007).
5. Collini, E. *et al. Nature* **463,** 644–647 (2010).
6. Sarovar, M., Ishizaki, A., Fleming, G. R. & Whaley, K. B. *Nature Phys.* **6,** 462–467 (2010).
7. Arndt, M., Juffmann, T. & Vedral, V. *HFSP J.* **3,** 386–400 (2009).

CANCER

Tumour stem cells switch sides

Tumour stem cells are proposed to be the source of tumour cells. It now emerges that they also give rise to the endothelial cells that line the tumour vasculature, mediating tumour growth and metastasis. **See Letters p.824 & p.829**

VICTORIA L. BAUTCH

To grow, solid tumours need a blood supply. They recruit new blood vessels mainly by inducing the sprouting of endothelial cells from external vessels and promoting the cells' migration into the tumour. This ability, called the angiogenic switch, is required for tumour cells to invade surrounding tissue and metastasize to distant sites — the deadly hallmarks of cancer[1]. In this issue, Wang *et al.*[2] and Ricci-Vitiani *et al.*[3] show that, in addition to recruiting vessels from outside, brain tumours produce endothelial cells for vessel formation from within.

Recent research in tumour biology has focused on two main concepts. According to the first concept — vasculogenic mimicry — some tumour cells take on certain characteristics of vascular endothelial cells and line the tumour's blood vessels[4]. The origin of such tumour cells is ill-defined: whereas one study[5] suggested that tumour stem cells show vasculogenic mimicry, it is generally thought that tumour cells in the immediate environment of the nascent vessel are co-opted for the purpose. The co-opted cells are thought to retain most of their tumour-cell characteristics while acquiring a limited number of endothelial-cell features.

The second concept — that some tumours originate from a tumour stem cell — has been controversial. According to this idea, tumour stem cells are both refractory to most traditional therapies and capable of regenerating the tumour following treatment. The deadly brain tumour glioblastoma is thought to arise from tumour stem cells[6].

Wang *et al.*[2] (page 829) and Ricci-Vitiani *et al.*[3] (page 824) now reveal data that are relevant to both concepts, and provide strong evidence that a proportion of the endothelial cells that contribute to blood vessels in glioblastoma originate from the tumour itself, having differentiated from tumour stem-like cells.

Both groups note that a subset of endothelial cells lining tumour vessels carry genetic abnormalities found in the tumour cells themselves (Fig. 1). For instance, a comparable proportion of a cell population expressing endothelial-cell markers and a population of neighbouring tumour cells harboured three or more copies of either the *EGFR* gene or other parts of chromosome 7. Such cell populations also shared a mutated version of the oncogene *p53*. Another indicator of the tumour origin of some tumour-vessel endothelial cells is that, as well as expressing characteristic endothelial-cell markers — such as von Willebrand factor and VE-cadherin — they expressed the non-endothelial, tumour marker GFAP.

The researchers also present evidence that tumour-derived endothelial cells arise from tumour stem-like cells. They find that a glioblastoma cell population that could differentiate into endothelial cells and form blood vessels *in vitro* was enriched in cells expressing the tumour-stem-cell marker CD133. Moreover, Wang and colleagues show that a clone of cells derived from a single tumour cell, which expressed CD133 but not VE-cadherin, was multipotent: *in vitro*, the cells differentiated into both neural cells (which eventually form tumour cells) and endothelial cells.

On being grafted into mice, these cells formed highly vascularized tumours. Moreover, even the progenitor cells from these tumours continued to form tumours and tumour-derived endothelial cells, suggesting that the multipotential characteristic had been maintained. Ricci-Vitiani *et al.* gained further insights by generating undifferentiated cell aggregates from human tumour-derived CD133-expressing cells and grafting them into mice. The internal vessels of the resulting tumours expressed human vascular markers,

whereas more external vessels carried mouse-specific endothelial-cell markers. What's more, the authors found human endothelial cells in tumour vessels linking to the mouse vessels and delivering blood to the tumour.

Wang *et al.*[2] suggest that the differentiation of tumour stem-like cells into endothelial cells might be mediated by signalling pathways involving two proteins — vascular endothelial growth factor (VEGF) and Notch. The authors propose that Notch regulates the initial differentiation of tumour stem-like cells to endothelial progenitor cells, whereas VEGF selectively affects the differentiation of endothelial progenitors to tumour-derived endothelial cells (Fig. 1).

Another team[7] has also investigated the source of cells contributing to tumour vessels, and has shown that tumour stem-like cells cultured from human glioma tumours form endothelial cells *in vitro*. The authors detected channels lined with tumour-derived cells in mice transplanted with human tumours — a process they classify as vasculogenic mimicry. However, their analysis of the original human tumours was limited to marker expression, and so they could draw no firm conclusion about the relationship between the tumour cells and the endothelial cells. Similarly, other groups[8,9] have presented evidence of genetic abnormalities common to tumour cells and endothelial cells, but their data did not distinguish among several potential mechanisms for the observations.

What is the functional significance of a tumour origin for vascular endothelium? To address this question, Ricci-Vitiani *et al.*[3] generated tumours in which the tumour-derived vessels were susceptible to drug-mediated destruction. Following drug treatment, these tumours were smaller than control tumours and had fewer blood vessels. This indicates that blood vessels derived from tumours are crucial for tumour survival.

The new work[2,3] also defines the relationship between a tumour and the blood vessels with which it interacts. If a dedicated compartment of some tumours provides a niche for stem cells that can give rise to functional blood vessels, there may be a less urgent need for tumour cells to undergo the angiogenic switch to recruit vessels, and stronger selective pressure on them to differentiate into endothelial cells.

Moreover, these observations challenge the assumption that tumour endothelial cells are normal cells, and therefore lack the genetic instability that may be the basis of drug resistance in tumour cells. Consistent with this suggestion, earlier studies[10,11] showed that tumour endothelial cells over-duplicate centrosomes — cellular organelles involved in cell division — and possess elevated levels of chromosome abnormalities. Moreover, there seems to be a link between increased activity of the signalling cascades that promote blood-vessel formation and chromosome abnormalities in endothelial

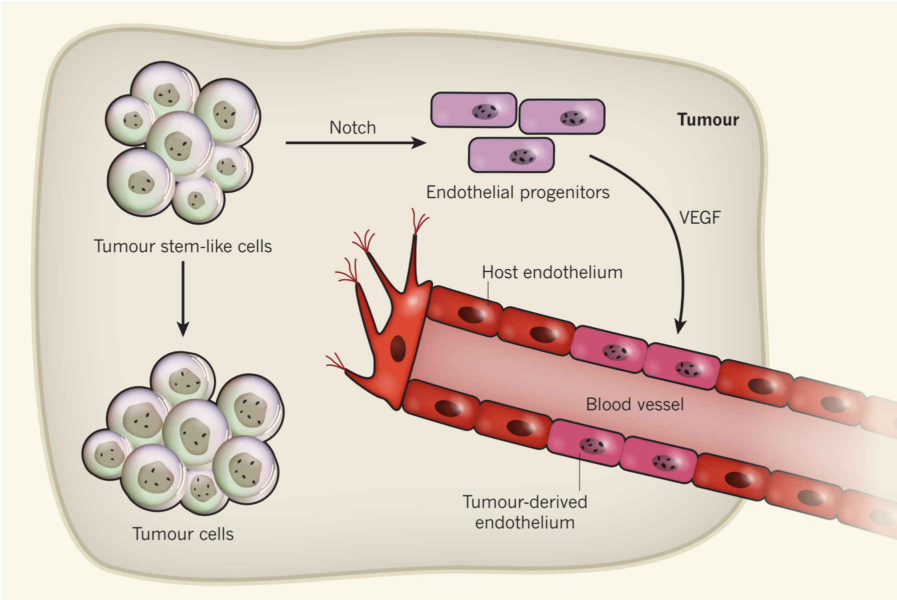


Figure 1 | Tumour stem-like cells are multipotent. Tumour-derived stem cells are thought to give rise to tumour cells. Wang *et al.*[2] and Ricci-Vitiani *et al.*[3] propose that a portion of the vascular endothelium that lines the tumour vessels in glioblastoma also arises from tumour stem-like cells. They show that the genetic abnormalities (dots) seen in the tumour cells are also present in endothelial cells isolated from the tumours. It seems that tumour stem-like cell differentiation to endothelial-cell progenitors occurs through Notch-mediated signalling, and that further differentiation of endothelial-cell progenitors into endothelial cells is mediated by the VEGF signalling pathway.

cells[12]. Tumour cells may therefore promote genetic instability in tumour endothelial cells through two distinct mechanisms: by giving rise to them directly, or by sending a signal to a nearby endothelial cell. Thus, not only the tumour compartment, but also genetically unstable tumour endothelial cells, may contribute to drug resistance.

Several compelling questions arise from the latest data[2,3]. First, how general is the differentiation of tumour stem-like cells into endothelial cells? Both studies focused on glioblastomas, and so the relevance of this pathway in other tumours of suspected stem-cell origin must also be determined. Other cell types of the underlying support tissue (stroma), such as fibroblasts, also play a part in tumour formation and progression. Do tumour stem cells contribute to these non-endothelial stromal lineages, and, if so, under what conditions?

It is also necessary to define the conditions that promote the differentiation of tumour stem-like cells to endothelial cells, and to determine the prevalence of this process within a given tumour environment. For example, does local shortage of oxygen trigger this differentiation? The present studies examine the molecular pathways that regulate the formation of tumour-derived endothelium at a superficial level. Defining the relevant mechanisms thoroughly is an essential prelude to the design of new therapies.

Finally, it will be crucial to determine how tumour-derived endothelial cells and vessels differ from their non-tumour counterparts in both morphology and function. Other studies[13] have reported that, when cultured, endothelial cells isolated from tumours exhibit some properties of stem cells, with the assumption that these properties were acquired by signals from the tumour environment. In light of the present work, an intriguing alternative possibility is that endothelium derived from tumour stem-like cells contributes to the observed cell characteristics. This work[2,3] therefore highlights yet another of the numerous ways in which tumours evade destruction: by contributing to their own support system. ■

Victoria L. Bautch *is in the Department of Biology, Lineberger Comprehensive Cancer Center, McAllister Heart Institute, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA.*
e-mail: bautch@med.unc.edu

1. Hanahan, D. & Folkman, J. *Cell* **86**, 353–364 (1996).
2. Wang, R. *et al. Nature* **468**, 829–833 (2010).
3. Ricci-Vitiani, L. *et al. Nature* **468**, 824–828 (2010).
4. Hendrix, M. J. C., Seftor, E. A., Hess, A. R. & Seftor, R. E. B. *Nature Rev. Cancer* **3**, 411–421 (2003).
5. El Hallani, S. *et al. Brain* **133**, 973–982 (2010).
6. Lathia, J. D., Venere, M., Rao, M. S. & Rich, J. N. *Stem Cell Rev. Rep.* doi:10.1007/s12105-010-9194-6 (2010).
7. Dong, J. *et al. Stem Cell Rev. Rep.* doi:10.1007/s12105-010-9169-7 (2010).
8. Streubel, B. *et al. N. Engl. J. Med.* **351**, 250–259 (2004).
9. Pezzolo, A. *et al. J. Clin. Oncol.* **25**, 376–383 (2007).
10. Hida, K. *et al. Cancer Res.* **64**, 8249–8255 (2004).
11. Akino, T. *et al. Am. J. Pathol.* **175**, 2657–2667 (2009).
12. Taylor, S. M. *et al. Blood* **116**, 3108–3117 (2010).
13. Dudley, A. C. *et al. Cancer Cell* **14**, 201–211 (2008).

50 YEARS AGO

The July issue of *Man* contains several articles of general interest … A. D. Lacaille illustrates a number of very large British Acheulean coups de poing, and a puzzling rock-carving from the Val Camonica is discussed by Dr. Anati of Paris. The site is near where the great glacial valley debouches on to the north Italian plain, and many rock-carvings there have been known for a long time. They include animals and humans treated in a conventional manner somewhat recalling the Copper Age paintings of Las Figuras in south-west Spain. The little group in question seems to indicate either a phallic or a ritual scene. The author suggests a date for this art group somewhere towards the start of the first millennium B.C. Is not this somewhat too early?
From *Nature* 10 December 1960.

100 YEARS AGO

The Anatomy of the Honey Bee. By R. E. Snodgrass. — In this modest pamphlet the author has given to entomologists an original, trustworthy, and excellently illustrated account of the structure of the honey bee … Many volumes have been written on the honey bee, yet no surprise can be felt that Mr. Snodgrass has been able to add new points to our knowledge and to correct errors in the work of his predecessors … He expresses scepticism as to certain positive statements that have been made on controverted details of physiology and reproduction; for example, "concerning the origin of the royal jelly or of any of the larval food paste … we do not know anything about it." There is a present-day tendency unduly to disparage the results obtained by former workers, and such a statement will strike many readers as extreme.
From *Nature* 8 December 1910.

COSMOLOGY

Hydrogen was not ionized abruptly

When and how the first stars and galaxies ionized the primordial hydrogen atoms that filled the early Universe is not known. Observations with a single radio antenna are opening a new window on the process. SEE LETTER P.796

JONATHAN PRITCHARD & ABRAHAM LOEB

Four hundred thousand years after the Big Bang, the Universe had cooled sufficiently for hydrogen atoms to form. Hundreds of millions of years later, the first stars and galaxies had produced ionizing ultraviolet radiation that broke the hydrogen atoms into their constituent electrons and protons. This process, termed reionization, marks a major cosmological phase transition. When and how rapid this transition was are important open questions[1]. On page 796 of this issue, Bowman and Rogers[2] implement a new technique that allows them to rule out models in which reionization occurs abruptly.

Their approach uses a simple radio antenna operating at low frequencies to measure the absolute radio intensity of the sky. Cosmic hydrogen atoms can emit or absorb light with a wavelength of 21 centimetres, a signal that is stretched (redshifted) on its way to Earth through the expansion of the Universe[3]. The redshifted 21-cm hydrogen signal, which falls within the radio regime, is expected to cut off at short, observed wavelengths that correspond to later times when the Universe was ionized. The authors' experiment to detect the global reionization step (EDGES) searches for the associated spectral step in the sky's intensity[4].

Our knowledge of the epoch of reionization is surprisingly limited. The lack of ultraviolet (UV) absorption by diffuse neutral hydrogen along the line of sight to the most distant quasars[5] (accreting black holes) indicates that the Universe is largely ionized at a redshift of less than about 6 — a billion years after the Big Bang. Yet observations of the cosmic microwave background[6] — radiation left over from the Big Bang — indicate that the Universe was filled with neutral hydrogen at much earlier times. Clearly, a transition must have occurred from a neutral to an ionized Universe, but even recent observations of high-redshift galaxies with the Hubble Space Telescope tell us little

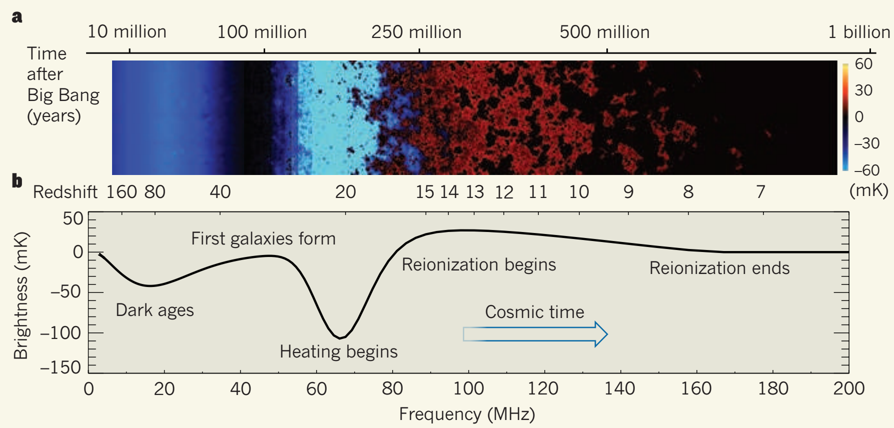


Figure 1 | The 21-centimetre cosmic hydrogen signal. **a**, Time evolution of fluctuations in the 21-cm brightness from just before the first stars formed through to the end of the reionization epoch. This evolution is pieced together from redshift slices through a simulated cosmic volume[9]. Coloration indicates the strength of the 21-cm brightness as it evolves through two absorption phases (purple and blue), separated by a period (black) where the excitation temperature of the 21-cm hydrogen transition decouples from the temperature of the hydrogen gas, before it transitions to emission (red) and finally disappears (black) owing to the ionization of the hydrogen gas. **b**, Expected evolution of the sky-averaged 21-cm brightness[8] from the 'dark ages' at redshift 200 to the end of reionization, sometime before redshift 6. The frequency structure within this redshift range is driven by several physical processes, including the formation of the first galaxies and the heating and ionization of the hydrogen gas. There is considerable uncertainty in the exact form of this signal, arising from the poorly understood properties of the first galaxies. Bowman and Rogers[2] study the final phase, in which the progressive ionization of the gas cuts off the signal.

about the galaxies that must have driven reionization[7].

Two major challenges for detecting the 21-cm signal involve foregrounds and calibration. The cosmic signal is dwarfed by radio emission from the Milky Way, as well as by terrestrial radio emission. Despite the favourable location of the EDGES experiment in the Australian outback, transmission from local radio and TV stations causes the loss of isolated regions of the spectrum. In addition, Galactic radio emission from energetic electrons spiralling in magnetic fields forms a spectrally smooth foreground that is one-thousand times brighter than the 21-cm signal. This smooth Galactic foreground can be fitted with a simple polynomial, and so removed, leaving the cosmic signal in the residuals. Unfortunately, this procedure removes much of the signal, potentially throwing the baby out with the bath water.

Another important limitation of the current experimental set-up is the absence of a method for calibrating the frequency response of the radio antenna. This necessitates fitting a combination of the foregrounds and the antenna's response. Given these limitations, it is impressive that the authors[2] are able to achieve residuals at the level of tens of millikelvin, comparable to the expected signal, and to place weak constraints on the duration of reionization.

Bowman and Rogers' technique allows them to rule out only models in which reionization occurs most abruptly — corresponding to a redshift interval of less than 0.1. As yet, the technique has had little effect on most models of the reionization epoch and the first galaxies. Figure 1 shows the expected evolution of the Universe as traced by emission or absorption of the 21-cm spectral line[8]. There is an initial absorption regime where the hydrogen gas is cooling through its cosmic expansion, and the excitation temperature of the 21-cm transition, which characterizes the relative populations of its two energy levels, is held equal to the gas temperature by collisions between hydrogen atoms. This absorption dies away as the gas gets diluted. Then the first stars form and emit UV photons that again set the excitation temperature of the 21-cm transition equal to the gas temperature, reinvigorating a second absorption trough. As these stars die, some of them produce black holes whose X-ray emission is expected to heat the gas to above the temperature of the cosmic microwave background, pushing the 21-cm signal into emission.

The authors[2] focus their efforts on this final phase, in which the signal is seen in emission and the progressive ionization of the diffuse hydrogen gas cuts off the signal, indicating the end of reionization. The same technique could ultimately be applied to detecting earlier periods for which our picture of the astrophysics is highly uncertain.

In the meantime, considerable time and money is being dedicated to the construction of low-frequency radio interferometers such as MWA, LOFAR and PAPER, which will target spatial fluctuations in the 21-cm signal (Fig. 1a). The EDGES experiment represents a cheaper method for measuring only the sky-averaged, broad-brush features in the evolution of the signal. Despite its limitations, it opens the possibility of an alternative experimental avenue that should be pursued in parallel to the more ambitious interferometers. Bowman and Rogers[2] have taken the first step on this journey, which will hopefully lead to new insights about the first stars and galaxies and the reionization epoch. ■

Jonathan Pritchard *and* **Abraham Loeb** *are in the Institute for Theory and Computation, Harvard University, Cambridge, Massachusetts 02138, USA.*
e-mails: jpritchard@cfa.harvard.edu; aloeb@cfa.harvard.edu

1. Loeb, A. *How Did the First Stars and Galaxies Form?* (Princeton Univ. Press, 2010).
2. Bowman, J. D. & Rogers, A. E. E. *Nature* **468,** 796–798 (2010).
3. Furlanetto, S. R., Oh, S. P. & Briggs, F. H. *Phys. Rep.* **433,** 181–301 (2006).
4. Shaver, P. A., Windhorst, R. A., Madau, P. & de Bruyn, A. G. *Astron. Astrophys.* **345,** 380–390 (1999).
5. Becker, R. H. *et al. Astron. J.* **122,** 2850–2857 (2001).
6. Komatsu, E. *et al. Astrophys. J. Suppl. Ser.* **180,** 330–376 (2009).
7. Robertson, B. E., Ellis, R. S., Dunlop, J. S., McLure, R. J. & Stark, D. P. *Nature* **468,** 49–55 (2010).
8. Pritchard, J. R. & Loeb, A. *Phys. Rev. D* **82,** 023006 (2010).
9. Santos, M. G. *et al. Astrophys. J.* **689,** 1–16 (2008).

VIROLOGY

One protein, many functions

The Lassa virus nucleoprotein coats the viral genome to make a template for RNA synthesis. A study shows that it also binds the 'cap' structure of cellular messenger RNAs and directs immune evasion using a novel mechanism. SEE ARTICLE P.779

FÉLIX A. REY

Lassa fever is a dreadful human haemorrhagic disease caused by the Lassa virus, a member of the Arenaviridae family[1]. The disease is prevalent in West Africa, causing 5,000 deaths each year and infecting hundreds of thousands more[2]. Arenaviruses are distributed worldwide and cause persistent infection in rodents, in which they generally don't cause disease. Humans become infected by exposure to material contaminated by infected mice, for example when the animals infiltrate food stores. The Lassa virus genome is a negative-sense, single-stranded RNA (nsRNA) molecule, and is coated by a nucleoprotein to form a nucleocapsid — a complex in which multiple copies of the nucleoprotein wrap around the genomic RNA, each one contacting a fixed number of nucleotides. In this issue (page 779), Qi *et al.*[3] report the crystal structure of the Lassa virus nucleoprotein, and reveal that it has a striking array of activities.

The nucleocapsids of nsRNA viruses serve as templates for the virus's polymerase enzyme (also known as the large or L protein), which replicates the genome to make new infectious particles. Qi and colleagues' crystal structure[3] shows that the Lassa virus nucleoprotein is made of two domains — an amino-terminal domain and a carboxy-terminal domain — with a positively charged groove in between, where the genomic RNA is expected to bind. This organization has been observed in all nsRNA viruses for which the nucleoprotein structure is known.

Before replication, the polymerase transcribes the genome into messenger RNA molecules to be translated into the viral proteins. Efficient translation of mRNAs by cellular ribosomes occurs if the mRNAs have a 'cap' structure at the 5′ end of the molecule. But arenaviruses, along with a subset of nsRNA viruses (those that have segmented genomes; Fig. 1, overleaf), cannot themselves cap mRNAs. They therefore steal caps from cellular mRNAs and transfer them to nascent viral transcripts, in a process known as cap snatching. Arenaviruses do this by cleaving off the 5′ end of cellular mRNAs using an 'endonuclease' activity that resides in the amino-terminal domain of the L protein[4,5], and then transferring the mRNA fragment to nascent transcripts.

Qi and colleagues' structure of the Lassa virus nucleoprotein shows that its amino-terminal domain has a cap-binding site, which holds the 5′ end of cellular mRNAs in place while the L protein cleaves off the rest. This additional function of the arenavirus nucleoprotein has not been observed in counterparts of the protein from any other virus family. The authors[3] show that when key residues in the cap-binding site are mutated, transcription is impaired.

Furthermore, the structure shows that the carboxy-terminal domain of the Lassa virus's nucleoprotein is folded in the same way as cellular 3′–5′ exonucleases — the enzymes that remove nucleotides one at a time from the 3′ end of RNA or DNA molecules, often completely degrading the nucleic-acid molecules in the process. Indeed, one of the closest structural homologues of the nucleoprotein's carboxy-terminal domain is the human DNA 3′–5′ exonuclease enzyme TREX1. What are the implications of this?

The detection of foreign nucleic acids to induce production of type I interferon (IFN) proteins is central to the innate antiviral defence of cells; misregulation of this system causes autoimmune problems. TREX1 is necessary for the degradation of single-stranded DNA derived from endogenous retroelements[6] (which constitute 90% of the approximately three million transposable elements in the human genome). Such single-stranded DNA accumulates in TREX1-deficient cells, inducing the IFN response and causing autoimmune disease. Qi *et al.* identified the amino acids of the 3′–5′ exonuclease active site of the Lassa virus nucleoprotein by superposition of their crystal structure[3] on that[7] of TREX1. Remarkably, these amino acids correspond to residues that were recently shown to have a critical role in the IFN-counteracting activity of the nucleoprotein of the lymphocytic choriomeningitis virus[6], the best-studied arenavirus. This suggests that the 3′–5′ exonuclease activity is the way by which arenavirus nucleoproteins inhibit IFN induction.

The authors verified[3] that the Lassa virus nucleoprotein does indeed degrade short RNA molecules similar to those generated as by-products during replication and transcription of the virus. They also showed that the wild-type nucleoprotein inhibits IFN production in virus-infected cells, whereas mutants devoid of exonuclease activity do not, even though they still undergo replication and transcription.

Why are so many unrelated activities concentrated in a single protein? The answer is probably related to the extreme compactness of arenavirus genomes, which code for only four proteins — fewer than in any of the other nsRNA virus families, and fewer than in any other human pathogenic virus. Of these four proteins, two of them (the nucleoprotein and the L protein) are present in all nsRNA viruses, and form the replicative foundation of these viruses. The structure of the Lassa virus nucleoprotein thus also provides further insight into the evolutionary history of nsRNA viruses, as described below.

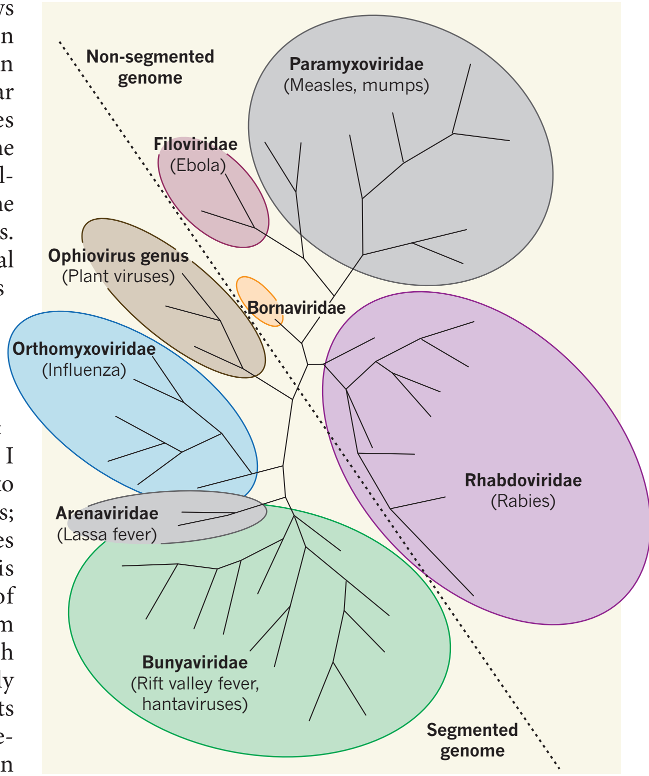


Figure 1 | Phylogeny of nsRNA viruses. This unrooted phylogenetic tree of the known nsRNA viruses (adapted from ref. 20) is overlaid with coloured ellipses representing the various virus families. The ophioviruses constitute a genus rather than a family. Representative human pathogens/diseases in each of the families are indicated in parentheses. The diagonal dashed line separates the nsRNA families that have a single RNA genomic molecule (non-segmented genomes) from those that have several genomic segments. The arenaviruses have two genomic segments, bunyaviruses have three, ophioviruses three or four (depending on the virus) and orthomyxoviruses six to eight. Qi *et al.*[3] report the crystal structure of the nucleoprotein from the Lassa virus (a member of the Arenaviridae family). The structure reveals an unexpected biological function of the nucleoprotein, and casts fresh light on the evolutionary history of nsRNA viruses.

The amino-acid sequence of the arenavirus L protein has the signature of RNA-dependent RNA polymerases (RdRps, enzymes that catalyse the replication of RNA from an RNA template). L proteins are found in all nsRNA viruses, with the exception of those of the Orthomyxoviridae family, in which the polymerase is split into three smaller polypeptides (PA, PB1, PB2) and functions as a heterotrimer containing these three proteins[8] — PA has the cap-snatching endonuclease site[9], PB1 acts as the catalytic RdRp[10] and PB2 has the cap-binding site[11]. Qi *et al.*[3] have now shown that, in arenaviruses, the cap-binding site resides in the nucleoprotein. The Bunyaviridae family of nsRNA viruses, meanwhile, have endonuclease activity in the amino-terminal domain of the L protein[12], but their cap-binding site has not yet been identified. Thus, the three families of nsRNA viruses that have segmented genomes and that have been studied in detail (Bunyaviridae, Orthomyxoviridae and Arenaviridae) share a cap-snatching strategy for genome transcription. This is not the case in the non-segmented viruses, in which the L protein has a capping activity.

The phylogenetic diagram shown in Figure 1 is based on the conserved RdRp modules from all nsRNA viruses, and shows how the different families cluster according to whether or not they have segmented genomes. Structural data show that the nucleoproteins from all non-segmented nsRNA viruses have evolutionarily conserved folds, but this isn't the case for the segmented ones. So, although the nucleoprotein structures of the bunyaviruses[13] and the orthomyxoviruses[14] both contain two domains (an amino-terminal and a carboxy-terminal domain, as seen for the Lassa virus nucleoprotein[3]), the individual folds of the domains are unrelated. By contrast, the amino- and carboxy-terminal domains of the nucleoproteins of the Bornaviridae[15], the Rhabdoviridae[16,17] and the Paramyxoviridae[18] (all of which have non-segmented genomes) have a conserved three-dimensional fold, suggesting a common ancestry, despite the absence of any detectable similarity in their amino-acid sequences.

Qi and colleagues' crystal structure[3] of an arenavirus nucleoprotein illuminates the protein's roles in the virus's cycle, while adding to our understanding of the evolutionary history of nsRNA viruses. It also highlights the fact that each family of nsRNA viruses seems to have developed different immune-defence strategies and shows that the arenaviruses' mechanism of immune evasion is a novel one. Considering that fatal infections by pathogenic arenaviruses — and chiefly by the Lassa virus[19] — are characterized by a generalized immune suppression, these new results have major implications for finding new ways to combat these diseases. ■

Félix A. Rey *is in the Structural Virology Unit, Department of Virology, Institut Pasteur, 25 rue du Dr Roux, Paris 75015, France.*
e-mail: rey@pasteur.fr

1. Buchmeier, M. J., De la Torre, J. C. & Peters, C. J. in *Fields Virology* Vol. 2 (eds Knipe, D. M. & Howley, P. M.) 1791–1828 (Lippincott Williams & Wilkins, 2007).
2. Ogbu, O., Ajuluchukwu, E. & Uneke, C. J. *J. Vector Borne Dis.* **44,** 1–11 (2007).
3. Qi, X. *et al. Nature* **468,** 779–783 (2010).
4. Lelke, M., Brunotte, L., Busch, C. & Günther, S. *J. Virol.* **84,** 1934–1944 (2010).
5. Morin, B. *et al. PLoS Pathog.* **6,** 1–11 (2010).
6. Stetson, D. B., Ko, J. S., Heidmann, T. & Medzhitov, R. *Cell* **134,** 587–598 (2008).

7. de Silva, U. *et al. J. Biol. Chem.* **282,** 10537–10543 (2007).
8. Ruigrok, R. W., Crepin, T., Hart, D. J. & Cusack, S. *Curr. Opin. Struct. Biol.* **20,** 104–113 (2010).
9. Dias, A. *et al. Nature* **458,** 914–918 (2009).
10. Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. *EMBO J.* **8,** 3867–3874 (1989).
11. Guilligay, D. *et al. Nature Struct. Mol. Biol.* **15,** 500–506 (2008).
12. Reguera, J., Weber, F. & Cusack, S. *PLoS Pathog.* **6,** e1001101 (2010).
13. Raymond, D. D., Piper, M. E., Gerrard, S. R. & Smith, J. L. *Proc. Natl Acad. Sci. USA* **107,** 11769–11774 (2010).
14. Ye, Q., Krug, R. M. & Tao, Y. J. *Nature* **444,** 1078–1082 (2006).
15. Rudolph, M. G. *et al. Structure* **11,** 1219–1226 (2003).
16. Albertini, A. A. *et al. Science* **313,** 360–363 (2006).
17. Green, T. J., Zhang, X., Wertz, G. W. & Luo, M. *Science* **313,** 357–360 (2006).
18. Tawar, R. G. *et al. Science* **326,** 1279–1283 (2009).
19. Baize, S. *et al. J. Virol.* **83,** 5890–5903 (2009).
20. Fauquet, C. *et al. Virus Taxonomy* (Elsevier, 2005).

SOLAR SYSTEM

Pluto is again a harbinger

New astronomical and laboratory data show that the abundances of the two dominant ices, nitrogen and methane, on the surfaces of the Solar System's two largest dwarf planets are surprisingly similar — raising fresh questions.

S. ALAN STERN

Combining state-of-the-art telescopic and ground-based laboratory data, Tegler *et al.*[1] have recently reported that the proportions of nitrogen (N_2) and methane (CH_4), the dominant surface ices on the two largest dwarf planets, Pluto and Eris, are surprisingly similar. More specifically, they found that the N_2 and CH_4 abundances on Eris are near 90% and 10%, respectively, and that those on Pluto are 97% and 3%. Intriguingly, these abundances are also similar to those on the dwarf planet and Kuiper-belt escapee Triton, which orbits Neptune.

Tegler and colleagues' results, published in the *Astrophysical Journal*, represent the first quantitative comparison of the abundances of volatile ices on the surface of any bodies beyond Neptune. They have significant implications for understanding Pluto and Eris, as well as the Kuiper belt, the disk-shaped region beyond Neptune's orbit where these two dwarf planets and other bodies reside (Fig. 1). The findings also provide reassurance that the detailed study planned for the Pluto system by NASA's New Horizons mission[2], which is now en route for a 2015 fly-by, will be of relevance to a broader suite of small planets common to the outer Solar System.

The discovery of Pluto by Clyde William Tombaugh in 1930 can be considered the technical discovery of the Kuiper belt. But the Kuiper belt's existence was firmly established only in the 1990s, with the discovery of additional bodies there[3,4]. Interestingly, a wide variety of attributes now known to be common to many large Kuiper-belt objects were first identified in studies of Pluto[4]. These include Pluto's rocky interior, its icy red surface, the presence of its satellites, its high orbital inclination and its resonant orbit with Neptune (Pluto's orbital

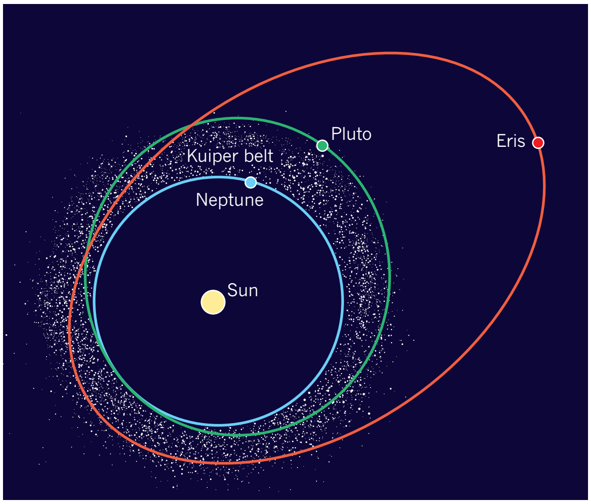


Figure 1 | Pluto, Eris and the Kuiper belt. Tegler and colleagues' demonstration[1] that Pluto and Eris have similar surface abundances of nitrogen and methane ices suggests that such abundances may be common, or at least not uncommon, among large objects in the Kuiper belt, the disk-shaped region beyond Neptune's orbit where the two dwarf planets reside. White dots represent objects in the classical Kuiper belt. Neither Centaurs (Kuiper-belt escapees) nor objects in the 'scattered belt' beyond Pluto's orbit are shown. Other large dwarf planets smaller than Pluto and Eris are also not shown.

period is in the precise ratio of 3/2 of Neptune's orbital period). As such, it is reasonable to refer to Pluto as the harbinger of the Kuiper belt and many of its key attributes.

The discoveries of CH_4 and N_2 ices on Pluto were reported in 1976 and 1992, respectively[5,6]. By discovering similar abundances of N_2 and CH_4 ices on Eris and Pluto, Tegler *et al.*[1] have demonstrated that Pluto's icy surface composition may be common — or at least not uncommon — among large Kuiper-belt worlds, thereby demonstrating another way in which Pluto seems to be a harbinger. Furthermore, because both N_2 and CH_4 create significant atmospheric vapour pressures at characteristic Kuiper-belt surface temperatures[7], an important implication of the authors' discovery[1] is that tenuous N_2–CH_4 atmospheres such as Pluto's (its atmospheric pressure is conceivably a few tens of microbars) may also be a common attribute among planets in the Kuiper belt.

Yet Tegler and colleagues' findings also raise new questions. A pivotal one is why some large Kuiper-belt worlds, such as Eris and Pluto, display N_2 and CH_4 on their surfaces, whereas others — even those that are similar to Eris and Pluto in both size and location in the Kuiper belt — display only H_2O ice on their surface, with no trace[8] of either N_2 or CH_4. A second, related, question concerns comparisons between the surface compositions of Pluto and Eris, and those of comets, which themselves derive from, and are thought to be the building blocks of, dwarf planets. Although the CH_4 fractions on Pluto and Eris are not unlike those seen in some comets[9], it is puzzling that they display so much N_2 on their surfaces when comets are apparently uniformly N_2 poor[9].

The ongoing rapid advance of ground-based astronomical facilities offers hope that, within this decade, such questions will be answered. The obvious route would be to apply Tegler and colleagues' methods — which involve both ground-based infrared spectroscopy and laboratory-based spectral studies of ice mixtures — to many more Kuiper-belt planets and smaller bodies.

Adding to the likelihood that such questions will be resolved in this decade are two important space missions now en route to their targets. One is the European Space Agency's flagship Rosetta comet orbiter, which will make the most detailed and comprehensive exploration ever imagined[10] of a comet (and Kuiper-belt escapee). Rosetta will arrive at its target, comet 67P/Churyumov–Gerasimenko, in mid-2014. Then, just one year later, NASA's New Horizons mission[2] will reconnoitre Pluto and all three of its known moons in exquisite detail.

Of particular relevance for surface-composition studies is the fact that both missions carry sensitive infrared mapping spectrometers. These spectrometers will, for the first time, reveal the distribution of N_2, CH_4 and many other compounds across the surface of a dwarf planet and search for them across a comet. What's more, they will, by dint of the close proximity of their spacecraft to the respective targets, also be able to look into the near-surface interiors of these representatives of comets and dwarf planets. This will be accomplished by examining the surface compositions of subsurface windows afforded by craters, fissures and exposed, vertically bedded layering where it is present on these bodies.

Tegler *et al.*[1] have revealed both compositional insight into, and commonalities among, the two largest planets of the Kuiper belt. It is up to future research teams, working with even more advanced facilities than those used by the authors, to address the questions that this discovery has raised, and to determine how much more diversity or commonality there is in surface composition among the planets of the Kuiper belt. ■

S. Alan Stern *is in the Space Science and Engineering Division, Southwest Research Institute, Boulder, Colorado 80302, USA.*
e-mail: alan@boulder.swri.edu

1. Tegler, S. C. *et al. Astrophys. J.* **725,** 1296–1305 (2010).
2. Stern, A. & Spencer, J. *Earth Moon Planets* **92,** 477–482 (2003).
3. Jewitt, D. *Annu. Rev. Earth Planet. Sci.* **27,** 287–312 (1999).
4. Davies, J. K. *Beyond Pluto: Exploring the Outer Limits of the Solar System* (Cambridge Univ. Press, 2001).
5. Cruikshank, D. P., Pilcher, C. B. & Morrison, D. *Science* **194,** 835–837 (1976).
6. Owen, T. C. *et al. Science* **261,** 745–748 (1993).
7. Washburn, E. W. *International Critical Tables of Numerical Data, Physics, Chemistry and Technology* (McGraw-Hill, 1933).
8. Pinilla-Alonso, N., Licandro, J., Gil-Hutton, R. & Brunetto, R. *Astron. Astrophys.* **468,** L25–L28 (2007).
9. Mumma, M. J., Weissman, P. R. & Stern, S. A. in *Protostars and Planets III* (eds Levy, E. H. & Lunine, J. I.) 1177–1252 (Univ. Arizona Press, 1993).
10. Verdant, M. & Schwehm, G. *ESA Bull.* **93,** 1–13 (1998).

GLACIOLOGY

Greenland's glacial basics

Sliding of the Greenland ice sheet is affected by the production of surface meltwater. A new theory shows that whether the result is a long-term speed-up or slow-down of ice motion depends on the variability in melt input. SEE LETTER P.803

MARTIN P. LÜTHI

The Greenland ice sheet is influenced by warming of its environment in three ways: higher melt rates at the surface, faster ice loss to the ocean, and faster (or maybe slower) sliding over the base. Melting is well understood, and big leaps have been made recently in understanding how glacier calving is influenced by warmer ocean currents. Our knowledge of sliding, however, lacks an essential factor — a universal relation linking sliding speed to the stress state at the glacier base, and to the main actor, pressurized subglacial water.

In a milestone study on page 803 of this issue[1], Schoof provides a unified theory of subglacial water drainage. He illustrates how the drainage system switches between different modes while adapting to the variable input of surface water, and why variability in water input, rather than the total water volume, drives ice-sheet acceleration.

The Swiss physicist and geologist Horace-Bénédict de Saussure suspected as early as 1779 that water drives glacier motion[2]. Many observations[3] confirm the more explicit statement that the pressure of subglacial water controls sliding processes. And because surface water from rain and melt finds its way to the glacier base, the subglacial water pressure varies within minutes — and with it ice velocity. On mountain glaciers, 20–80% of the displacement observed on the surface is attributable to sliding processes at the base, the rest being due to ice deformation[4]. The extreme value of more than 99% has been observed on West Antarctic ice streams, whereas at the other end of the scale are extended areas in which the ice sheets are frozen to the bed. The only direct measurement made in Greenland (60%) is not representative of this ice sheet, because the drill site concerned was near a fast-paced outlet glacier[5].

How future changes in precipitation and melt will affect subglacial water pressure, and therefore ice velocity, is crucial for predicting the future evolution of the Greenland ice sheet. If more meltwater input increases basal motion, more ice will be transported to lower elevations, leading to drawdown of the ice-sheet surface, and so to further melting. Rapid ice loss at rates exceeding previous estimates would thus be conceivable[6]. But current

understanding of subglacial hydrology and basal motion do not support the idea that ice-sheet sliding increases with the amount of water production during the summer melt season. Schoof's model calculations[1], based on the unified description of subglacial hydrology shown in detail in the online Supplementary Information for his paper, suggest that proponents of both ideas are right — at least partly.

The unification offered by Schoof concerns two major types of subglacial drainage system: a distributed system of linked cavities[7], and a system of channels within the ice or along the ice–bed boundary[8]. The distributed system consists of cavities that form in the lee of protrusions in the glacier bed, and that grow with water pressure and with sliding speed. This system operates at high water pressure caused by inefficient drainage of water. On the other hand, channels become enlarged through melting of surrounding ice, and can grow without limit because dissipative heat production increases with water flux at high pressure gradients. Big channels therefore operate at lower pressure and grow at the expense of smaller channels, which leads to the evolution of an arborescent structure similar to an arterial network.

Both types of drainage system adapt to changes in water input within hours to days, and collapse by the inward creeping motion of the ice when the water pressure drops below the ice-overburden pressure. The switch from a linked cavity system to an arborescent channel network is beautifully illustrated in Schoof's animations, which are part of his Supplementary Information.

To understand the consequences of the switch in drainage-system configuration, consider the effect of a sudden increase in water discharge on ice-sheet motion. If additional surface water reaches a distributed system, water pressure increases, large patches of ice are separated from the bed, and the contact area between ice and bed is diminished. The result is less friction, and faster sliding of ice over the bed, which again leads to growth of cavities. Sudden increase in water input has a similar initial effect on a channelized drainage system. But higher pressure gradients immediately lead to higher discharge, and to a rapid adaptation to the new water input by channel enlargement. Once fully adapted, the bigger channel dimensions even lead to lower water pressure, which drives more water from the surroundings to the channel. The net effect is better coupling to the bed and slower sliding motion.

Schoof's results[1] highlight an important effect that has been largely ignored until recently. Surface water is mainly released in pulses — through daily melt, rain or breakthrough drainage from lakes on the ice surface — which are usually shorter than the timescale of channel enlargement. Such brief pressure pulses drive water from channels into the surrounding distributed system, with accelerated sliding being the result. To sum up, releasing a certain amount of water at a steady rate leads to initial acceleration, and subsequent deceleration, whereas release in pulses leads to episodic speed-up events that add up to larger displacements.

J. GULLY

Figure 1 | Exploring a subglacial artery. Glaciologists from the University Centre in Svalbard exploit a rare circumstance in investigating a major drainage channel under a glacier (the Rieperbreen glacier on Svalbard, an archipelago in the Arctic). Subglacial channels such as this are efficient pathways for the escape of highly pressurized water that would otherwise lead to fast sliding motion of the glacier over the bed. They are transient features that form during the melt season — when surface water penetrates to the glacier base — and that slowly collapse in winter.

The main obstacle to successful application of the proposed drainage-system model to big ice sheets is the nearly complete lack of field data from which to determine model parameters. The remoteness of the subglacial environment makes conditions under an ice sheet notoriously difficult to observe (except in rare situations, as shown in Fig. 1). Measuring basal motion and water-pressure variations in 'wet-based' parts of the Greenland ice sheet requires drilling through 500–2,500 metres of ice — an expensive endeavour that has been successful in only four locations[5,9,10]. Driven by the pressing need to predict the future behaviour of the Greenland ice sheet, several current drilling projects aim to observe the basal drainage system in action.

Schoof's paper[1] constitutes a notable advance in understanding subglacial processes. However, development of a complete theory will require a fuller knowledge of several other factors. The motion of a glacier over its substrate is a spatially distributed phenomenon that involves frictional processes at a wide spectrum of scales — solid friction between ice and bedrock; granular friction within sediments ranging in size from silt to boulders; friction between sediment and bedrock; and drag on the ice when it flows around obstacles. And there is another hydrological effect to be taken into account: diffusion of highly pressurized subglacial water through sediments influences their rheology in a time-dependent manner. Even after 50 years of ingenious experimental and theoretical advances, much remains to be done, in terms of both fieldwork and theory. ■

Martin P. Lüthi *is in the VAW Glaciology Group, ETH Zürich, 8092 Zürich, Switzerland.*
e-mail: luethi@vaw.baug.ethz.ch

1. Schoof, C. *Nature* **468,** 803–806 (2010).
2. de Saussure, H.-B. *Voyage dans les Alpes* (4 vols) (Fauche, 1779–1796).
3. Clarke, G. K. C. *Annu. Rev. Earth Planet. Sci.* **33,** 247–276 (2005).
4. Cuffey, K. M. & Paterson, W. S. B. *The Physics of Glaciers* 4th edn (Elsevier, 2010).
5. Lüthi, M. P., Funk, M., Iken, A., Gogineni, S. & Truffer, M. *J. Glaciol.* **48,** 369–385 (2002).
6. Parizek, B. R. & Alley, R. B. *Quat. Sci. Rev.* **23,** 1013–1027 (2004).
7. Kamb, B. *Rev. Geophys. Space Phys.* **8,** 673–728 (1970).
8. Röthlisberger, H. *J. Glaciol.* **11,** 177–203 (1972).
9. Thomsen, H. H. & Olesen, O. B. *Rap. Grønl. Geol. Unders.* **152,** 36–38 (1991).
10. Iken, A., Echelmeyer, K., Harrison, W. D. & Funk, M. *J. Glaciol.* **39,** 15–25 (1993).

doi:10.1038/nature09605

Cap binding and immune evasion revealed by Lassa nucleoprotein structure

Xiaoxuan Qi[1], Shuiyun Lan[2], Wenjian Wang[3], Lisa McLay Schelde[2], Haohao Dong[1], Gregor D. Wallat[1], Hinh Ly[2], Yuying Liang[2] & Changjiang Dong[1]

Lassa virus, the causative agent of Lassa fever, causes thousands of deaths annually and is a biological threat agent, for which there is no vaccine and limited therapy. The nucleoprotein (NP) of Lassa virus has essential roles in viral RNA synthesis and immune suppression, the molecular mechanisms of which are poorly understood. Here we report the crystal structure of Lassa virus NP at 1.80 Å resolution, which reveals amino (N)- and carboxy (C)-terminal domains with structures unlike any of the reported viral NPs. The N domain folds into a novel structure with a deep cavity for binding the m7GpppN cap structure that is required for viral RNA transcription, whereas the C domain contains 3′–5′ exoribonuclease activity involved in suppressing interferon induction. To our knowledge this is the first X-ray crystal structure solved for an arenaviral NP, which reveals its unexpected functions and indicates unique mechanisms in cap binding and immune evasion. These findings provide great potential for vaccine and drug development.

Several arenaviruses, including Lassa virus (LASV), can cause severe viral haemorrhagic fevers in humans with high morbidity and mortality, to which there is no vaccine and limited treatment[1–4]. These pathogenic arenaviruses are public health threats and potential biological threat agents. LASV, like other arenaviruses, is a single-stranded ambisense RNA virus with two genomic RNA segments encoding four genes[1]. The NP encapsidates viral genomic RNAs into ribonucleoprotein (RNP) complexes and is required for both RNA replication and transcription[5–7]. Like bunyaviruses and orthomyxoviruses, arenaviruses snatch the cap structure of cellular mRNAs to use as primers to initiate viral transcription, the exact mechanism of which is unknown. The cap-snatching mechanism of arenaviruses seems to be unique, as evidenced by the cytoplasmic localization and the much shorter 5′ non-templated mRNA sequences[8–11]. Severe arenavirus infections including lethal Lassa cases are associated with a generalized immune suppression in the infected hosts[12–18], the exact mechanism of which is unclear but is thought to involve NP's ability to suppress the induction of type I interferon (IFN)[19,20]. To address the functional mechanisms of NP in viral RNA synthesis and host immune suppression, we set out to determine the crystal structure for LASV NP, knowledge derived from which can be extended to other arenavirus NP proteins, as all known arenaviral NP proteins share high sequence identity (Supplementary Fig. 1).

Structure determination

The full-length 569-residue LASV NP protein (Josiah strain) was expressed and purified as a recombinant MBP fusion protein in *Escherichia coli* as described in Methods. The purified protein exists mainly in two forms, with a majority in trimeric and some in hexameric form. Both forms bind random RNAs, which are longer and more abundant in the hexamers than in the trimers, a feature that is similar to known NPs from negative-strand RNA viruses[21–24]. We attempted to crystallize both forms, but only the trimeric NP formed crystals. The crystals showed heavy twining with a twin fraction of ~0.43 and the reflection intensity statistic $|E^2\text{-}1|$ 0.681/0.681. Initial phases were obtained in a space group of *P*321 using the multiple wavelength anomalous diffraction (MAD) with Samarium derivative. The true space group was *P*3 with three subunits in an asymmetric unit. The structure was refined to a resolution of 1.80 Å with de-twining. The crystal structures do not contain RNA, indicating that only RNA-free NP was able to form crystals. The final structural model of the native LASV NP has an R_{factor} of 0.18 and an R_{free} of 0.20. Data collection, phasing and refinement statistics are provided in Supplementary Table 1.

Overall structure of LASV NP protein

In the NP protomer structure, 514 residues of the 569-residue LASV NP protein were built into the model (Fig. 1a). The electron densities for residues 1–6, 147–157, 339–363, 518–521, 562–569 were not well defined. LASV NP protomer, like other viral NPs[7–10], is composed of the N- and the C-terminal domains, but neither domain shows structural similarity to any known viral NPs (Supplementary Table 2). The large N domain (residues 7–338) consists mainly of α-helices and coils, whereas the C domain (residues 364–561) forms a typical α/β/α sandwich architecture (Supplementary Text 1). In the trimeric form, three subunits lie in a head-to-tail orientation to form a ring-shaped structure with a three-fold symmetry (Fig. 1b and Supplementary Fig. 2). Surface rendering reveals a deep cavity located near the bottom of the N domain and a large cavity at the top of the C domain (Fig. 1c, d), which are the cap-binding site and the 3′–5′ exoribonuclease active site (see below), respectively. The interface area between the subunits is 455 $Å^2$, representing 1.9% of total surface area of a subunit (23,343 $Å^2$). The central hole of the trimeric structure is 23 Å in diameter, whereas the head ring is 98 Å and the body ring is 118 Å (Supplementary Fig. 2).

LASV NP is a 3′–5′ exoribonuclease

A Dali search (http://ekhidna.biocenter.helsinki.fi/dali_server) identified several structures similar to the C domain of NP, including several known 3′–5′ exonucleases/exoribonucleases in bacteria and humans (for example, human TREX1) (Supplementary Text 2), all of which belong to the DEDDH subfamily of the DEDD (DnaQ) superfamily[25–27]. The

[1]Biomedical Sciences Research Complex, School of Chemistry, University of St Andrews, North Haugh, St Andrews, Fife KY16 9ST, UK. [2]Department of Pathology and Laboratory Medicine, Emory University School of Medicine, 615 Michael St, Atlanta, Georgia 30322, USA. [3]Laboratory of Department of Surgery, The First Affiliated Hospital, Sun Yat-Sen University, 58 Zhongshan Road II, Guangzhou, Guangdong 510080, China.

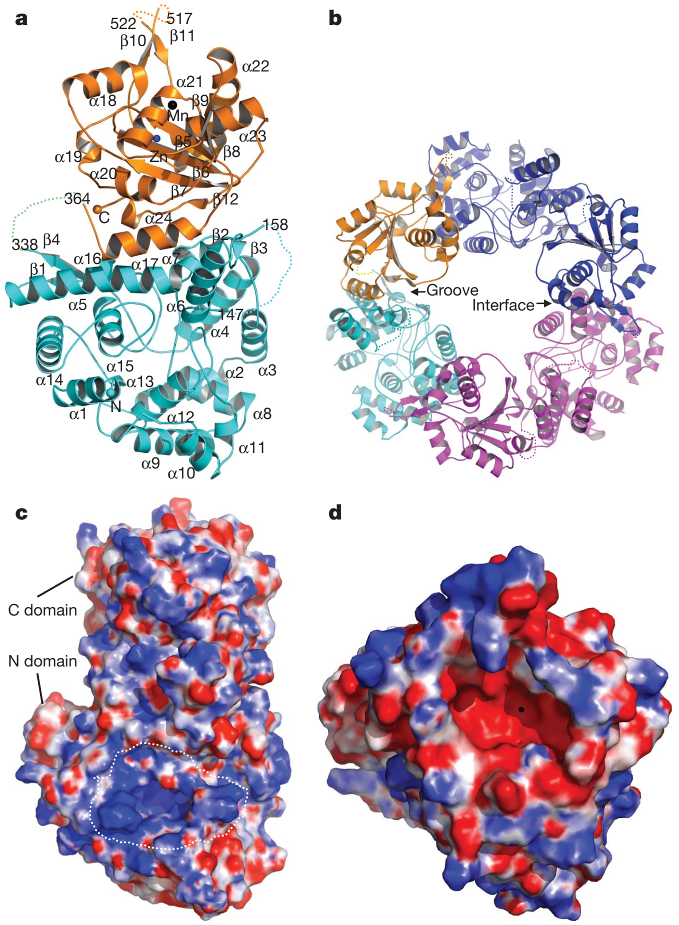


Figure 1 | The crystal structure of LASV NP protein. **a**, Cartoon diagram of the LASV NP protomer. The N domain is in cyan with the cyan sphere indicating the N terminus; the C domain is in orange with the orange sphere indicating the C terminus. The black sphere shows Mn^{2+}, whereas the blue sphere shows Zn^{2+}. The dotted lines represent the disordered loops. **b**, The ring-shaped structure of LASV NP trimer. The first protomer is coloured as in **a**, the second protomer is in blue and the third is in magenta. The groove and the interface are indicated by arrows. **c**, Electrostatic surface potential map of the NP protomer. The entrance of the cap-binding cavity is shown as a white dotted circle. The blue area represents positively charged residues and the red area represents negatively charged residues. **d**, Electrostatic surface potential map of the 3′–5′ exoribonuclease cavity. The black sphere represents Mn^{2+}.

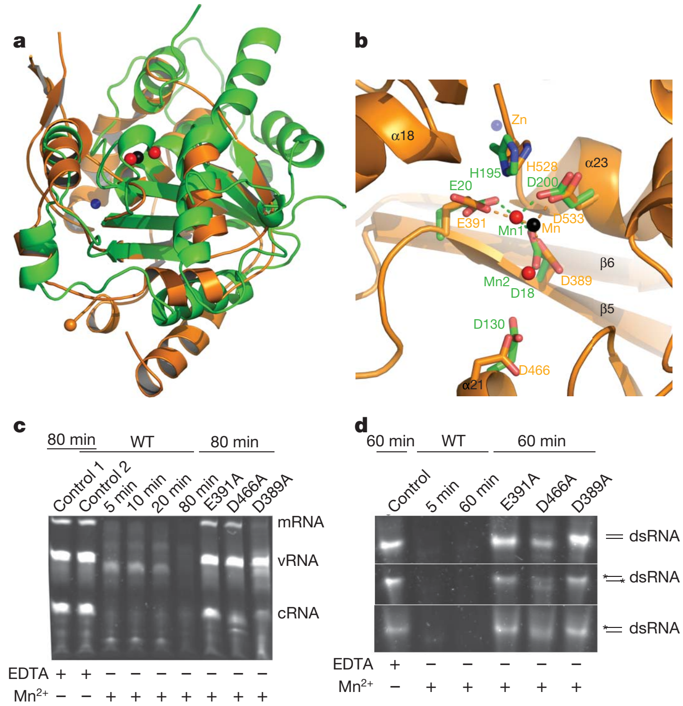


Figure 2 | The C domain of LASV NP is a 3′–5′ exoribonuclease. **a**, Superimposition of the C domain (orange) with human TREX1 protein (green) reveals a high degree of similarity between the two structures. Mn^{2+} is in black in LASV NP and red in TREX1; Zn^{2+} is in blue. **b**, The exonuclease catalytic residues of LASV NP and TREX1 are located in identical positions, and are shown in orange for NP and green for TREX1. **c**, The exoribonuclease activities of the wild-type (WT) and mutant LASV NP with different ssRNAs as substrates. Control 1 contains 10 mM EDTA and no NP. Control 2 contains 10 mM EDTA and NP. **d**, Comparison of the wild-type and NP catalytic mutants in degrading the dsRNA substrates, the 5′-hydroxyl dsRNA (top), double 5′-triphosphorylated dsRNA (middle), and the single 5′-triphosphorylated dsRNA (bottom).

human TREX1 structure shows two Mn^{2+} cations in the active site[27]. We identified one Mn^{2+} in each subunit of LASV NP by crystal fluorescent scanning, but could not identify the second Mn^{2+}, possibly because it was not well ordered in the absence of the RNA substrate. The C domain of NP superimposes well with the portion of TREX1 that coordinates the Mn^{2+} cations (Fig. 2a), in particular the β5, β6, β7, β8 and β9 strands of NP completely overlap with the central β-sheets of TREX1. The putative exonuclease catalytic residues D389, E391, D466, D533 and H528 are absolutely conserved in all known arenavirus NP proteins and are located at identical positions as in the TREX1 active cavity (Fig. 2b). Taken together, the structural evidence indicates that LASV NP is a new member of the DEDD 3′–5′ exonuclease superfamily.

We conducted *in vitro* assays to characterize the 3′–5′ exonuclease activity of the wild-type LASV NP, as well as NP mutants at putative catalytic sites. We showed that the wild-type protein, in its trimeric or hexameric form, could digest both DNA and RNA substrates (Supplementary Figs 3 and 4). As divalent cations are essential for exonuclease activity[25], we determined what divalent cation was most effective for NP exonuclease to digest various single-stranded RNA (ssRNA) species that are based on the NP gene in the viral genomic sense (60 nucleotides, vRNA), complementary antigenomic sense (30 nucleotides, cRNA), or in capped mRNA form (126 nucleotides, mRNA) (Methods). We showed the order of efficiency as $Mn^{2+}>Co^{2+}>Mg^{2+}>Ca^{2+}>Zn^{2+}>Fe^{2+}>Ni^{2+}>Cu^{2+}$ (Supplementary Fig. 5). Wild-type NP could cleave various ssRNA species efficiently (Fig. 2c), regardless of whether they contained a hydroxyl (5′OH) group, triphosphate (5′ppp), or a cap at the 5′ termini (Methods). In contrast, the NP catalytic mutants (D389A, E391A and D466A) showed markedly reduced RNase activity (Fig. 2c and Supplementary Fig. 4). In addition, we showed that wild-type NP, but not its catalytic mutants, could digest cellular RNA substrates *in vitro* with a preference towards short RNA species over long ones (for example, 18s rRNA versus β-globin mRNA, the large versus small fragments in the RNA ladder) (Supplementary Figs 6 and 7). We also demonstrated that wild-type NP, but not its catalytic mutants (D389A, E391A and D466A), can efficiently degrade various dsRNA molecules with 5′-hydroxyl (5′OH), single 5′-triphosphorylate (5′ppp/5′OH) and double 5′-triphosphorylate (5′ppp/5′ppp), as well as the long dsRNA mimic poly(I:C) (Fig. 2d and Supplementary Fig. 8).

Fluorescence scanning analysis identified a zinc ion in the NP structure, despite the fact that no typical zinc finger motif was predicted from the amino acid sequence and that no zinc compounds were used during the purification and crystallization processes. Although the residues C506, C529, H509 and E399 that coordinate the zinc ion are not of the typical zinc-binding motif[28], they appear to adopt a zinc finger fold in structure[28,29]. The CCHE zinc-binding site is located in the C domain near the 3′–5′ exonuclease active site (Supplementary Fig. 9). We speculate that zinc binding may be required to

stabilize the structure of the C domain and/or contribute to the substrate binding and specificity of the exonuclease activity[28,29]. A highly positively charged groove located between the N and C domains is predicted as the genomic RNA-binding site (Supplementary Fig. 10). An *in vitro* assay confirmed that RNAs are bound within the purified NP oligomers and protected from its intrinsic exonuclease activity (Supplementary Figs 10 and 11, Supplementary Text 3 and Methods).

Exonuclease and immune evasion

To determine whether the exoribonuclease activity is important for the transcriptional function of NP, we generated alanine substitution at five putative catalytic sites, D389A, E391A, D466A, D533A and H528A, in the mammalian cell expression vectors of either native or Myc-tagged NP gene, and examined the activity of each mutant in transcribing the LASV minigenome RNA that encodes a *Renilla* luciferase (RLuc) reporter gene[20] (Methods). As shown in Fig. 3a, each NP mutant expressed comparable protein levels to the wild type, and led to similar folds of increase in RLuc activity, indicating that these mutations did not alter the overall structure (Supplementary Text 4 and Supplementary Fig. 12) or affect the basic function of NP in mediating viral RNA transcription.

We next examined whether the exoribonuclease activity is required for NP's function in the suppression of IFN[19,20]. As expected, wild-type NP strongly inhibited Sendai-virus-induced IFN-β activation by a promoter assay (Methods), whereas all the catalytic mutants D389A, E391A, D466A, D533A and H528A showed a complete loss of function at a low level of transfected expression vectors (10 ng) and showed various levels of deficiency at higher levels (Fig. 3b and Supplementary Fig. 13). Our results confirm a previous study showing that the D389 residue of LASV NP, as well as its corresponding residue D382 in the prototypic arenavirus lymphocytic choriomeningitis virus (LCMV), is required for IFN suppression but not for viral RNA transcription[30], and may help to explain the loss of IFN suppression for Tacaribe virus NP (Supplementary Fig. 14 and Supplementary Text 5). In summary, these data provide strong genetic evidence for an important role of the NP exoribonuclease activity in suppressing the IFN induction.

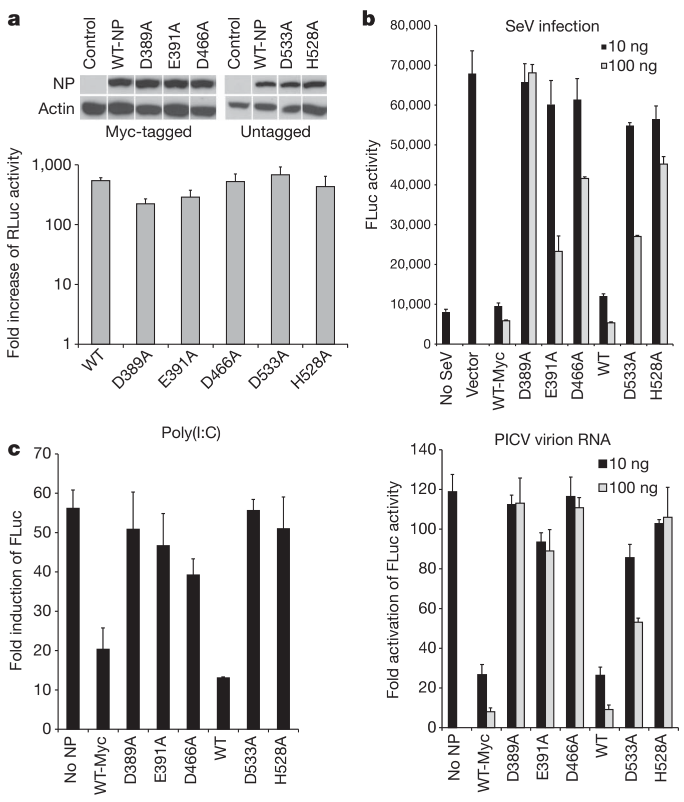


Figure 3 | The exonuclease activity of NP is important for blocking the IFN induction. Results shown are the average ($n = 3$) with error bars indicating the standard deviations. **a**, The NP catalytic mutants were expressed at similar levels to the wild type in mammalian cells and had similar transcriptional activities in the LASV minigenome assay. **b**, The NP catalytic mutants were defective in suppressing the Sendai-virus (SeV)-induced IFN induction by a LUC-based IFN-β promoter assay. **c**, The NP catalytic mutants were defective in suppressing the IFN production induced by the immunostimulatory RNAs poly(I:C) and Pichinde-virion-associated RNAs.

Viral infections are usually detected by the cellular pattern-recognition receptors (PRRs) such as toll-like receptors (TLRs) and cytosolic RNA sensors, retinoid-acid-inducible gene-I-like helicase (RIG-I) and melanoma differentiation-associated protein 5 (MDA5), which recognize the pathogen-associated molecular patterns (PAMP) RNA ligands and initiate signalling pathways to induce the production of type I IFNs[31,32]. We hypothesize that NP prevents the virus-induced IFN induction by degrading the PAMP RNA ligands that otherwise would trigger the viral sensors in the cells.

We examined whether the NP RNase function is essential for suppressing the IFN production induced by the immunostimulatory RNAs, that is, poly(I:C) and the virion RNAs extracted from Pichinde virus, which is a prototypic arenavirus[33]. We found that whereas wild-type NP efficiently inhibited the IFN-β activation induced by poly(I:C) or by Pichinde-virion-associated RNAs, none of the five catalytic mutants (D389A, E391A, D466A, D533A and H528A) exhibited any suppressive activity (Fig. 3c). Similar results had been reported for LCMV NP[34].

We have shown that the NP exoribonuclease activity is essential for suppressing both viral-infection-induced and immunostimulatory-RNA-induced IFN production. A good example of exonuclease-meditated suppression of IFN production has been demonstrated for human TREX1 protein, which degrades small ssDNAs and dsDNAs accumulated during cellular apoptosis. Failure to clear these DNA fragments by TREX1 natural mutants leads to the activation of cellular DNA receptors to trigger a persistent production of IFNs that contributes to human autoimmune diseases[27,35–38]. How does the NP RNase activity function in suppressing the virus-induced IFN production? A simplistic but reasonable model is that the NP RNase activity is able to remove viral PAMP RNAs that are otherwise recognized by the cellular PRRs. Although we have shown that LASV NP protein can degrade various RNA templates *in vitro*, we believe that the NP RNase activity must be highly regulated *in vivo*, as NP does not cause a generalized nonspecific RNA degradation process of cellular or viral RNAs in the cells (Supplementary Text 6 and Supplementary Fig. 15). We propose that the NP RNase activity in the cells is restricted to viral PAMP RNAs through a yet-to-be characterized regulatory mechanism. A recent publication has shown a direct protein–protein interaction of NP with RIG-I and MDA5 (ref. 34), which may be one possible mechanism for the specific nuclease activity of NP against these PRR-associated PAMP RNAs.

LASV NP is a cap-binding protein

The N domain adopts a completely novel fold not found in the Dali server. To identify the cap-binding residues in the deep cavity of the N domain, we attempted to soak and perform co-crystallization of LASV NP with m7GpppG, triphosphorylated, diphosphorylated or monophosphorylated ribonucleotides (Methods). We could observe the clear density for the triphosphate and partial density for uridine (Supplementary Fig. 16) from the triphosphorylated ribonucleotide complex structures. We also visualized the structure of NP in complex with dTTP with a clear original $F_o - F_c$ electron density contoured at 2.5σ for dTTP (Fig. 4a). The triphosphate group of dTTP was bound in the middle of the cavity in an identical manner as that of UTP (Supplementary Fig. 16), in which it was anchored by salt bonds formed with the side chains of the conserved residues K309, R300, R323 and K253. In the deep end of the cavity, thymidine occupied a hydrophobic pocket that is composed of residues F176, W164, L172,

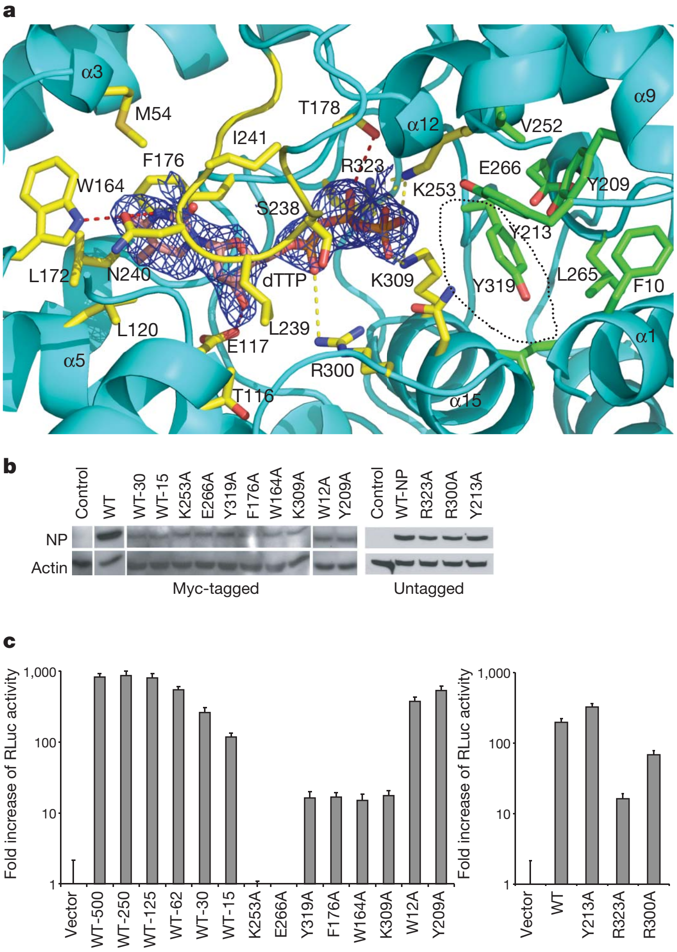


Figure 4 | The cap-binding residues and their roles in viral RNA transcription. Results shown are the average ($n = 3$) with error bars indicating the standard deviations. **a**, A cap analogue dTTP is bound within the deep cavity of the N domain of LASV NP. Original $F_o - F_c$ map for the dTTP in blue contoured at 2.5σ. The F176 and W164 or L172 (L120) residues form a typical cap-binding sandwich structure. The middle cavity binds the triphosphate moiety and the hydrophobic cavity entrance can accommodate the second base of the cap structure. The carbon atoms are in pink for the dTTP, in yellow for the deep cavity residues and in green for the cavity entrance residues. **b**, The NP mutants were expressed at similar levels as the wild type at 15–30 ng plasmid (WT-15, WT-30) in the transfected mammalian cells. **c**, Mutational analyses of the residues within the cap-binding cavity for the transcriptional activity using the LASV minigenome assay.

M54, L120, L239 and I241. We propose that this dTTP-binding pocket is the binding site for the cap structure m7GTP and that the residues located within the pocket may have to change conformation to accommodate the cap moiety. Although the N domain of NP is not structurally similar to any of the cap-binding proteins (Supplementary Table 3), its hydrophobic thymidine-binding pocket shares common features for cap binding[39,40]. Moreover, the NP cap-binding cavity has a unique feature in that its entrance contains another hydrophobic region that is composed of the hydrophobic residues Y319, Y209, Y213, L265 and the acidic residue E266, which can potentially act as the binding site for the second base of the m7GpppN (where N represents G, C, U or A) cap structure. The entrance of the cap-binding cavity has an oval shape with a diameter of 9–13 Å, which is a perfect fit for the single-stranded mRNA. We propose that a loop composed of residues K236 to S242 serves as a 'gate' for the capped template (primer) binding and that the entire structure of m7GpppN, including the cap m7G, the triphosphate, and at least one more nucleotide, is embedded within the deep cavity. This binding feature is unlike other known cap-binding proteins, in which only the m7G caps are locked in between the sandwich, whereas the rest of the RNA molecule is exposed[39,40].

To characterize the role of the cap-binding residues in viral RNA transcription, we examined a panel of NP mutants with alanine substitution of residues located inside and at the entrance of the cavity and that are conserved among all known arenaviruses for their ability to mediate the cap-dependent viral RNA transcription using the LASV minigenome replicon assay (Methods). Wild-type NP (with or without Myc tag) produced up to a 1,000-fold increase in RLuc reporter activity over a control reaction, and more than 100-fold increase even when expressed at a low level (15 ng of transfected NP plasmid DNAs). All mutant proteins were expressed at similar levels as the wild type transfected with 15–30 ng of plasmid (Fig. 4b). Compared to the wild type, the K253A and E266A mutants completely lost the RNA transcription activity, and the Y319A, F176A, W164A, K309A and R323A mutants showed significantly decreased activity (Fig. 4c). R300A had a minor effect, whereas W12A and Y209A had no effect. It is worth noting that none of these mutants was found to impact the NP function in the suppression of IFN (Supplementary Fig. 17). These functional data correlate well with the proposed cap-binding function of some of these conserved residues.

The unique cap-binding feature of LASV NP, in that the entire cap structure m7GpppN is buried within the cavity, has significant implications in understanding the distinctive cap-snatching mechanism of arenaviruses. Once NP binds and protects the 5′ cap m7GpppN, the rest of the mRNA molecule located outside of the cavity may be susceptible to viral and/or host exonuclease-mediated degradation and/or to endonuclease-mediated cleavage (Supplementary Fig. 9). This may help to explain the relatively short (1–4 nucleotides) 5′ non-templated sequences in arenavirus mRNAs[1,8,9]. However, individual mutation of the NP exonuclease catalytic sites did not show any defect in viral cap-dependent RNA transcription (Fig. 3a), indicating that the NP exonuclease activity is not essential (required) for generating the capped primers. It is worth noting that we did not identify an influenza polymerase PA-like endonuclease structural motif[41,42] within LASV NP structure (Supplementary Table 4). Instead, recent studies indicated that the LASV L polymerase protein contains an endonuclease domain in its N terminus that is crucial for the cap-dependent viral RNA transcription[43,44].

Conclusion

Our structural analysis and functional assays have demonstrated that the C domain of LASV NP contains 3′–5′ exoribonuclease activity that is required for suppressing IFN-β induction. We have provided evidence to suggest that the NP RNase activity is highly regulated in cells and proposed a novel mechanism by which the NP RNase activity may specifically remove the viral PAMP RNA ligands to suppress the production of IFN. Another important feature of LASV NP protein is that its N domain contains a deep cavity to bind and shield the entire m7GpppN cap structure, which is distinct from other known cap-binding proteins, and has shed light on the unique cap-snatching mechanism of arenaviruses. In addition, we have also identified an unusual zinc-binding site and the viral RNA-binding groove in the LASV NP structure. Taken together, these findings reveal several new and potentially vulnerable targets on NP for the development of antivirals and effective vaccines to combat LASV and other pathogenic arenaviruses that can cause severe haemorrhagic fever diseases in humans.

METHODS SUMMARY

The crystals were grown using the sitting-drop technique, and the native structure was determined with the MAD data. All the NP mutations were generated using the QuikChange site-directed mutagenesis kit (Stratagene) and confirmed by DNA sequencing. The RNA synthesis assays used the LASV minigenome (MG) system, and the Sendai-virus-induced IFN-β activation assay was conducted as described[45]. The immunostimulatory RNA-induced IFN-β activation assay was

conducted by transfecting HEK293 cells with the IFN-β-LUC promoter construct and either wild-type or mutant NP construct, followed by Lipofectamine-2000-mediated transfection of poly(I:C) or Pichinde-virion-isolated RNAs. Activation of the IFN-β promoter was quantified by measuring the LUC activity.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 May; accepted 25 October 2010.
Published online 17 November 2010.

1. Buchmeier, M. J., De La Torre, J. C. & Peters, C. J. in *Fields Virology* Vol. 2 (eds Knipe, D. M. & Howley, P. M.) 1791–1827 (Lippincott Williams & Wilkins, 2007).
2. Delgado, S. *et al.* Chapare virus, a newly discovered arenavirus isolated from a fatal hemorrhagic fever case in Bolivia. *PLoS Pathog.* **4,** e1000047 (2008).
3. Briese, T. *et al.* Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS Pathog.* **5,** e1000455 (2009).
4. Khan, S. H. *et al.* New opportunities for field research on the pathogenesis and treatment of Lassa fever. *Antiviral Res.* **78,** 103–115 (2008).
5. Hass, M., Golnitz, U., Muller, S., Becker-Ziaja, B. & Gunther, S. Replicon system for Lassa virus. *J. Virol.* **78,** 13793–13803 (2004).
6. Pinschewer, D. D., Perez, M. & de la Torre, J. C. Role of the virus nucleoprotein in the regulation of lymphocytic choriomeningitis virus transcription and RNA replication. *J. Virol.* **77,** 3882–3887 (2003).
7. Lopez, N., Jacamo, R. & Franze-Fernandez, M. T. Transcription and RNA replication of tacaribe virus genome and antigenome analogs require N and L proteins: Z protein is an inhibitor of these processes. *J. Virol.* **75,** 12241–12251 (2001).
8. Polyak, S. J., Zheng, S. & Harnish, D. G. 5′ termini of Pichinde arenavirus S RNAs and mRNAs contain nontemplated nucleotides. *J. Virol.* **69,** 3211–3215 (1995).
9. Meyer, B. J. & Southern, P. J. Concurrent sequence analysis of 5′ and 3′ RNA termini by intramolecular circularization reveals 5′ nontemplated bases and 3′ terminal heterogeneity for lymphocytic choriomeningitis virus mRNAs. *J. Virol.* **67,** 2621–2627 (1993).
10. Jin, H. & Elliott, R. M. Characterization of Bunyamwera virus S RNA that is transcribed and replicated by the L protein expressed from recombinant vaccinia virus. *J. Virol.* **67,** 1396–1404 (1993).
11. Plotch, S. J., Bouloy, M., Ulmanen, I. & Krug, R. M. A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **23,** 847–858 (1981).
12. Edington, G. M. & White, H. A. The pathology of Lassa fever. *Trans. R. Soc. Trop. Med. Hyg.* **66,** 381–389 (1972).
13. Fisher-Hoch, S., McCormick, J. B., Sasso, D. & Craven, R. B. Hematologic dysfunction in Lassa fever. *J. Med. Virol.* **26,** 127–135 (1988).
14. McCormick, J. B. & Fisher-Hoch, S. P. Lassa fever. *Curr. Top. Microbiol. Immunol.* **262,** 75–109 (2002).
15. Baize, S. *et al.* Early and strong immune responses are associated with control of viral replication and recovery in lassa virus-infected cynomolgus monkeys. *J. Virol.* **83,** 5890–5903 (2009).
16. Baize, S. *et al.* Lassa virus infection of human dendritic cells and macrophages is productive but fails to activate cells. *J. Immunol.* **172,** 2861–2869 (2004).
17. Mahanty, S. *et al.* Cutting edge: impairment of dendritic cells and adaptive immunity by Ebola and Lassa viruses. *J. Immunol.* **170,** 2797–2801 (2003).
18. Muller, S., Geffers, R. & Gunther, S. Analysis of gene expression in Lassa virus-infected HuH-7 cells. *J. Gen. Virol.* **88,** 1568–1575 (2007).
19. Martinez-Sobrido, L., Giannakas, P., Cubitt, B., Garcia-Sastre, A. & de la Torre, J. C. Differential inhibition of type I interferon induction by arenavirus nucleoproteins. *J. Virol.* **81,** 12696–12703 (2007).
20. Martinez-Sobrido, L., Zuniga, E. I., Rosario, D., Garcia-Sastre, A. & de la Torre, J. C. Inhibition of the type I interferon response by the nucleoprotein of the prototypic arenavirus lymphocytic choriomeningitis virus. *J. Virol.* **80,** 9192–9199 (2006).
21. Green, T. J., Zhang, X., Wertz, G. W. & Luo, M. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* **313,** 357–360 (2006).
22. Tawar, R. G. *et al.* Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. *Science* **326,** 1279–1283 (2009).
23. Ye, Q., Krug, R. M. & Tao, Y. J. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* **444,** 1078–1082 (2006).
24. Albertini, A. A. *et al.* Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* **313,** 360–363 (2006).
25. Cisneros, G. A. *et al.* Reaction mechanism of the epsilon subunit of *E. coli* DNA polymerase III: insights into active site metal coordination and catalytically significant residues. *J. Am. Chem. Soc.* **131,** 1550–1556 (2009).
26. Zuo, Y. *et al.* Crystal structure of RNase T, an exoribonuclease involved in tRNA maturation and end turnover. *Structure* **15,** 417–428 (2007).
27. de Silva, U. *et al.* The crystal structure of TREX1 explains the 3′ nucleotide specificity and reveals a polyproline II helix for protein partnering. *J. Biol. Chem.* **282,** 10537–10543 (2007).
28. Hall, T. M. Multiple modes of RNA recognition by zinc finger proteins. *Curr. Opin. Struct. Biol.* **15,** 367–373 (2005).
29. Matthews, J. M. & Sunde, M. Zinc fingers–folds for many occasions. *IUBMB Life* **54,** 351–355 (2002).
30. Martinez-Sobrido, L. *et al.* Identification of amino acid residues critical for the anti-interferon activity of the nucleoprotein of the prototypic arenavirus lymphocytic choriomeningitis virus. *J. Virol.* **83,** 11330–11340 (2009).
31. Kawai, T. & Akira, S. Innate immune recognition of viral infection. *Nature Immunol.* **7,** 131–137 (2006).
32. McCartney, S. A. & Colonna, M. Viral sensors: diversity in pathogen recognition. *Immunol. Rev.* **227,** 87–94 (2009).
33. Lan, S. *et al.* Development of infectious clones for virulent and avirulent pichinde viruses: a model virus to study arenavirus-induced hemorrhagic fevers. *J. Virol.* **83,** 6357–6362 (2009).
34. Zhou, S. *et al.* Induction and inhibition of type I interferon responses by distinct components of lymphocytic choriomeningitis virus. *J. Virol.* **84,** 9452–9462 (2010).
35. Crow, Y. J. & Rehwinkel, J. Aicardi-Goutieres syndrome and related phenotypes: linking nucleic acid metabolism with autoimmunity. *Hum. Mol. Genet.* **18,** R130–R136 (2009).
36. Lee-Kirsch, M. A. *et al.* Mutations in the gene encoding the 3′-5′ DNA exonuclease TREX1 are associated with systemic lupus erythematosus. *Nature Genet.* **39,** 1065–1067 (2007).
37. Lehtinen, D. A., Harvey, S., Mulcahy, M. J., Hollis, T. & Perrino, F. W. The TREX1 double-stranded DNA degradation activity is defective in dominant mutations associated with autoimmune disease. *J. Biol. Chem.* **283,** 31649–31656 (2008).
38. Stetson, D. B., Ko, J. S., Heidmann, T. & Medzhitov, R. Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell* **134,** 587–598 (2008).
39. Guilligay, D. *et al.* The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature Struct. Mol. Biol.* **15,** 500–506 (2008).
40. Fechter, P. & Brownlee, G. G. Recognition of mRNA cap structures by viral and cellular proteins. *J. Gen. Virol.* **86,** 1239–1249 (2005).
41. Yuan, P. *et al.* Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. *Nature* **458,** 909–913 (2009).
42. Dias, A. *et al.* The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* **458,** 914–918 (2009).
43. Lelke, M., Brunotte, L., Busch, C. & Gunther, S. An N-terminal region of Lassa virus L protein plays a critical role in transcription but not replication of the virus genome. *J. Virol.* **84,** 1934–1944 (2010).
44. Morin, B. *et al.* The N-terminal domain of the arenavirus L protein is an RNA endonuclease essential in mRNA transcription. *PLoS Pathog.* **6,** e1001038 (2010).
45. Lan, S., McLay, L., Aronson, J., Ly, H. & Liang, Y. Genome comparison of virulent and avirulent strains of the Pichinde arenavirus. *Arch. Virol.* **153,** 1241–1250 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements C.D. wishes to thank J. Naismith for his continuing support, encouragement and advice; R. M. Elliott for providing the human β-globin containing plasmid phRL-CMV and his critical reading of the manuscript; H. Liu for discussions; and L. Major for the pLou3 plasmid. C.D. is a Wellcome trust career development fellow (083501/Z/07/Z). We thank the staff at IO2 and IO3 beam stations of Diamond light sources for their assistance with data collection. This work was supported in part by funds from the Southeast Regional Center of Excellence for Emerging Infections and Biodefense (5-U54-AI-057157-06), the pilot component of the U19 grant (5-U19-AI057266-07), and the Emory University Research Committee (URC) to Y.L. and H.L.; a seed grant from the Emory DDRDC (DK64399) and a research scholar grant from the American Cancer Society (RSG-06-162-01-GMC) to H.L.; NIH grant R01AI083409 to Y.L.; and NIH grant AI067704 to T. G. Parslow and Y.L. We thank K. Curtis and T. W. Geisbert for providing us with non-infectious genomic RNA samples of LASV, and J. Aronson for the PICV virus.

Author Contributions X.Q., S.L., W.W., L.M.S., H.D. and G.D.W. performed experiments. C.D., Y.L. and H.L. conceived the idea for the study, performed some of the assays, participated in the analysis and interpretation of the data, and wrote the manuscript. All authors contributed to the final version of the manuscript.

Author Information Atomic coordinates and structure factors for the structures have been deposited in the Protein Data Bank under accession codes 3MWP for the native, 3MWT for the manganese complex, 3MX2 for the dTTP complex and 3MX5 for the UTP complex. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.D. (cd26@st-andrews.ac.uk), Y.L. (yliang5@emory.edu) or H.L. (hly@emory.edu).

METHODS

Protein expression and purification. The full-length LASV NP gene (Josiah strain) was cloned into the pMAL-c2X-derived pLou3 plasmid, downstream of the TEV cleavage site following the MBP gene. This construct, encoding the N-terminal MBP tagged NP protein, was transformed into Rosetta cells (Novagen). After IPTG induction at a final concentration of 0.03 mM overnight at 20 °C, the cells were harvested by centrifugation at 8,000 r.p.m. for 20 min and suspended in TEN buffer (20 mM Tris, pH 7.5; 0.2 M NaCl, 10% glycerol, 1 mM EDTA) with protease inhibitors (Roche), 1 μM DNase (Sigma) and 1 mM phenylmethylsulphonyl fluoride (Sigma). After cells were lysed by a cell disruptor (Constant System Ltd), the cell lysates were collected by centrifugation at 20,000 r.p.m. for 30 min and applied on an amylose column. The column was washed with >10-column volumes of the sample buffer. The MBP–NP fusion protein was eluted with the TEN buffer containing 10 mM maltose. The MBP–NP fusion protein was then cleaved by Tev proteinase. The MBP portion was removed through two amylose columns, and the NP protein was purified to homogeneity by gel filtration column. Trypsin digestion coupled with mass spectroscopy confirmed that the purified LASV NP protein was homogenous (data not shown), with a final concentration of 7 mg ml^{-1}.

Crystallization and data collection. A Cartesian robot (Genomic solutions) was used to screen for optimal crystallization conditions. The native crystals were obtained in 0.2 M $LiCl_2$ and 20% PEG3350 in 1 week at 20 °C. To obtain the NP complex with m7GpppG, m7GTP, or m7GDP, the NP protein was incubated with individual compound at a concentration of 2 mM for 30 min on ice and the crystallization conditions were screened. The NP complex with other triphosphorylated, diphosphorylated or monophosphorylated nucleotides were formed by incubating the NP protein with 50 mM of the respective compounds for 30 min on ice and the crystallization conditions were screened. The crystallization conditions were optimized until the resolution of the data was better than 2.5 Å. All crystals grew in 0.2 M KCl or 0.2 M LiCl and 14–22% PEG3350. The NP complexed with manganese ion was obtained by crystallizing the NP protein in 0.2 M $MnCl_2$, 25% PEG3350 followed by soaking the crystals in 0.2 M $NaCl_2$, 20% PEG3350 and 15% glycerol three times for 15 min each. The presence of the manganese and the zinc ions was confirmed in all the crystals by fluorescence scanning at the Diamond light sources UK. All the crystals were protected by cryoprotectants that contain 15% to 20% glycerol in the crystallization conditions before data collection in IO2 or IO3 at the Diamond light sources UK. The Samarium derivative crystals were obtained by soaking the crystals overnight in 100 mM Samarium acetate, 0.2 M LiCl and 16% PEG3350, and was protected in a cryoprotectant of 0.2 M LiCl, 16% PEG3350 and 20% glycerol. The Samarium derivative MAD data were collected at a wavelength of 1.83 Å for peak data, 1.84 Å for inflection data and 1.45 Å for remote data from a single crystal. All the data were indexed, integrated and scaled by HKL2000 or Mosflm and Scale.

Structure determination. The crystals were heavily twinned with a twining fraction of 0.43. The initial phases were obtained from a space group of *P*321 using the MAD data and SOLVE[46]. The initial model was built using RESOLVE[46], Buccaneer and Coot[47]. It was found that the true space group of the crystals was *P*3 during the structure refinement. The structures were refined using REFMAC5[48], and the water molecules were added into the structure by ARP/wARP[49]. The $F_o - F_c$ maps for ligands (dTTP, UTP, zinc and manganese) were calculated before any ligand was added into the structures. The structures were de-twinned at last using REFMAC5, and the structures were evaluated using Molprobity[50].

***In vitro* RNA synthesis.** The 30-nucleotide cRNA (sense) sequence 5′-CUGGGCUUACCUAUUCUCAGCUGAUGACCC-3′ was derived from the LASV NP (Josiah strain) S segment (nucleotides 2186–2215 in antigenomic orientation) and chemically synthesized by Eurogentic. The 30-nucleotide vRNA (in genomic orientation) sequence 5′-GGGUCAUCAGCUGAGAAUAGGUAAGCCCAG-3′ was complementary to the cRNA. The cRNA (30 nucleotides) was used as one of the three substrates for 3′–5′ exoribonuclease assay. To obtain the blunted dsRNA, both cRNA and vRNA oligonucleotides were dissolved into 0.1 M NaCl, 1 mM EDTA and 0.1 M Tris pH 8.0 at the final concentration of 200 mM, and an equal amount of the two oligonucleotides was mixed together and annealed in a thermocycler as follows: 95 °C for 3 min, 68 °C for 1 min and then 4 °C.

The 5′-triphosphorylated vRNA was generated by *in vitro* transcription of the partial dsDNA template formed by the T7 promoter sequence 5′-AATTTAATACGACTCACTATAGG-3′ and the reverse complement of the T7 promoter sequence and of the LASV (Josiah strain) S segment (nucleotides 2186–2215) 5′-CTGGGCTTACCTATTCTCAGCTGATGACCCTATAGTGAGTCGTATTAAATT-3′ using the T7 MEGAshortscript kit following the manufacturer's instructions (Ambion). A similar strategy was use to generate the 32-nucleotide triphosphorylated cRNA with the T7 primer and LASV (Josiah strain) S segment (nucleotides 2186–2213) 5′-GGGTCATCAGCTGAGAATAGGTAAGCCCAGCCTATAGTGAGTCGTATTAAATT-3′. A similar strategy was used to generate the 60-nucleotide vRNA corresponding to LASV (Josiah strain) S segment (nucleotides 2186–2213), using the partial dsDNA template formed by 5′-AATTTAATACGACTCACTATAGG-3′ and 5′-GTAAATCCCTGCAGTCGGCAGGGTTTACCGCTGGGCTTACCTATTCTCAGCTGATGACCCTATAGTGAGTCGTATTAAATT-3′ as a template. To generate the doubly 5′-triphosphorylated dsRNA, equal amounts of the triphosphorylated 5′ppp-vRNA and 5′ppp-cRNA were annealed *in vitro*. To make the singly 5′-triphosphorylated dsRNA, equal amounts of the *in vitro* synthesized 32-nucleotide 5′ppp-vRNA and the chemically synthesized 30-nucleotide unphosphorylated cRNA were annealed *in vitro*. The human 18S rRNA fragment (128 nucleotides) was generated by a T7 RNA polymerase-directed *in vitro* RNA synthesis reaction, using the pTRI-RNA 18S control plasmid (Ambion), following the manufacturer's instruction.

To synthesize the capped viral mRNA transcripts corresponding to nucleotides 992–1117 of the LASV NP gene, the DNA template was PCR amplified from the NP expression plasmid with a forward primer 5′-AATTTAATACGACTCACTATAGGGAAAACACTGTCGTTGATCTGGAATC-3′ (underlined are T7 promoter sequences) and a reverse primer 5′-GGGTCATCAGCTGAGAATAGGTAAGCCCAGCGG-3′, and subjected to *in vitro* RNA synthesis using the mMESSAGE mMACHINE T7 Ultra kit (Ambion) following the manufacturer's instruction, except that no poly(A) tail was added.

A plasmid phRL-CMV that encodes the T7 promoter (T7p)-directed human β-globin gene was provided by R. Elliott and G. Blakqori. The T7p-globin DNA fragment was purified by agarose electrophoresis after digestion of the phRL-CMV plasmid with HindIII and SmaI. The capped human globin mRNA transcripts were generated using the T7p-globin fragment as a template and the mMESSAGE mMACHINE T7 Ultra kit from Ambion, and the poly(A) tail was added following the manufacturer's instruction.

The ssRNA markers (perfect RNA markers, 0.1–1 kb) were purchased from Novagen. The low molecular mass ssRNA marker (10–100 nucleotides) was purchased from USB. The dsRNA ladder (21–500 bp) was purchased from New England Biolabs.

***In vitro* 3′–5′ exoribonuclease assays.** The *in vitro* 3′–5′ exoribonuclease assays were carried out in 10 μl of the reaction solution containing 0.3 M NaCl, 10% glycerol, 20 mM Tris pH 7.5, 10 mM $MnCl_2$, 7 μg of either wild-type or mutant NP proteins, and 8 units of the RNaseIN inhibitor (Promega), in the presence of various substrate(s), at 37 °C for 60–100 min. The control reactions included all but $MnCl_2$, which was substituted by 20 mM EDTA. All the reactions, each in triplicate, were stopped by the addition of EDTA to a final concentration of 20 mM. The samples were mixed with equal volumes of RNA loading buffer (Ambion), heated at 95 °C for 3 min, cooled on ice for 5 min, and separated in 15% or 6% urea-polyacrylamide gel, or 2% agarose gel. The gels were stained in 0.05% ethidium bromide for 25 min, visualized using the 2UV transilluminator (UVP).

The luciferase-based assay to quantify virus-induced and immunostimulatory RNA-induced interferon-β activation. The Sendai-virus-induced IFN-β activation assay was conducted as described previously[45]. In brief, 293T cells were co-transfected using calcium phosphate with 100 ng of a vector that expresses the firefly luciferase (FLuc) reporter gene from a known functional promoter sequence of the IFN-β gene (pIFNβ-LUC), variable amounts of either wild-type or mutant LASV NP vectors, and 50 ng of a β-gal-expressing plasmid for transfection normalization. At 24 h after transfection, cells were infected with Sendai virus (at multiplicity of infection = 1) to induce IFN-β expression. At 24 h after infection, cell lysates were prepared for luciferase and β-gal assays. FLuc activities were normalized by the β-gal values. Each transfection was conducted in triplicate and repeated in at least two independent experiments.

To determine whether NP can suppress the immunostimulatory RNA-induced IFN production, HEK293 cells were transfected with pIFNβ-LUC, variable amounts of either wild-type or mutant LASV NP vectors, and a β-gal-expressing plasmid for transfection normalization. Eighteen hours later, cells were transfected with either 1 μg of poly(I:C) or 250 ng of Pichinde virion RNA by lipofectamine 2000. Luciferase activity was determined at 18 h after the immunostimulatory RNA transfection and normalized by the β-gal activity.

Pichinde virion RNA preparation. Pichinde viruses were purified by 20% sucrose gradient ultracentrifugation at 50,000*g* for 2 h. Virus RNA was extracted with RNABee (Tel Test) according to the manufacturer's protocol.

LASV minigenome (MG) transcription assay. The full-length LASV L and NP genes (Josiah strain) were cloned into the pCAGGS vector for expression in mammalian cells. The LASV MG construct contains the T7 promoter-directed LASV S-segment-like sequences that include all the important cis-acting elements required for viral RNA synthesis (5′ UTR, intergenic region and 3′ UTR) and encode a *Renilla* luciferase (RLuc) gene in place of the viral NP coding sequence. This LASV-based LUC-encoding minigenome (MG) RNA was transcribed *in vitro* by the T7 MEGAScript kit (Ambion) and transfected into 293T cells, together with the LASV L expression plasmid, and wild-type or mutant NP expression plasmid. A β-gal expression vector was included in each

transfection to normalize for cell transfection efficiency. LUC activity was determined at 24 h after transfection, normalized by β-gal activity, and shown as fold increase over a control sample that lacked the L expression plasmid. Each reaction was conducted in triplicate and in at least two independent experiments.

46. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55,** 849–861 (1999).
47. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
48. Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D* **60,** 2184–2195 (2004).
49. Cohen, S. X. *et al.* Towards complete validated models in the next generation of ARP/wARP. *Acta Crystallogr. D* **60,** 2222–2229 (2004).
50. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35,** W375–W383 (2007).

ARTICLE

doi:10.1038/nature09516

Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA

Nicholas J. Reiter[1], Amy Osterman[1], Alfredo Torres-Larios[1]†, Kerren K. Swinger[1]†, Tao Pan[2] & Alfonso Mondragón[1]

Ribonuclease (RNase) P is the universal ribozyme responsible for 5′-end tRNA processing. We report the crystal structure of the *Thermotoga maritima* RNase P holoenzyme in complex with tRNA^{Phe}. The 154 kDa complex consists of a large catalytic RNA (P RNA), a small protein cofactor and a mature tRNA. The structure shows that RNA–RNA recognition occurs through shape complementarity, specific intermolecular contacts and base-pairing interactions. Soaks with a pre-tRNA 5′ leader sequence with and without metal help to identify the 5′ substrate path and potential catalytic metal ions. The protein binds on top of a universally conserved structural module in P RNA and interacts with the leader, but not with the mature tRNA. The active site is composed of phosphate backbone moieties, a universally conserved uridine nucleobase, and at least two catalytically important metal ions. The active site structure and conserved RNase P–tRNA contacts suggest a universal mechanism of catalysis by RNase P.

Ribonuclease P (RNase P) is a ribonucleoprotein complex responsible for processing many different RNA molecules in the cell (for recent reviews, see refs 1–3). It is found in almost all organisms and is composed of one essential RNA subunit and one or more protein subunits. The RNA component is responsible for catalysis and can process RNA *in vitro* in the absence of protein, albeit with reduced efficiency[4]. The discovery that the RNA component is the catalytic moiety[4] helped cement the notion that RNA can be directly involved in catalysis. RNase P is considered a remnant of an ancient RNA-based world and an example of an RNA-based catalyst with many features in common with protein-based catalysts.

RNase P recognizes its substrate in *trans* and is a multiple turnover enzyme. The preferred substrate is pre-tRNA and recognition involves features distant from the cleavage site, such as the TΨC loop of the tRNA acceptor stem[5]. RNA cleavage requires divalent metals[4,6,7], yet the chemical mechanism and the location of the active site remain largely undefined as well as the exact role of the protein components. In the case of bacterial RNase P, the single essential protein improves the reaction rate by two to three orders of magnitude[8,9], helps to stabilize the active P RNA fold[8,10], binds the 5′ leader region of the pre-tRNA substrate[11,12], and assists in product release[13].

Structural studies of the RNA component reveal a two domain (S- and C-domains) molecule formed by single and coaxial stems linked together by a variety of tertiary interactions[14–17], including five conserved regions I to V (CR-I to CR-V) of P RNA that are common to all organisms[18]. These conserved regions cluster into two areas, one involved in substrate recognition and the other forming the active site scaffold[19].

Here we present the crystal structure of *Thermotoga maritima* RNase P holoenzyme in complex with mature tRNA^{Phe}, and also the structure of the complex in the presence of a post-cleavage tRNA leader. The two structures help answer key questions about the mechanism of this crucial ribozyme with implications for a broader understanding of the general mechanisms of RNA–RNA based recognition and catalysis.

Structure determination

The components of the complex were purified separately and assembled by mixing and heating before crystallization (see Methods). The pre-tRNA was processed into mature tRNA and hence the structure represents a ribozyme–product complex. To promote crystal formation, two interaction modules[20] were introduced, which had a modest effect on catalytic activity (Supplementary Fig. 1 and Supplementary Table 1). The crystals diffract anisotropically to 3.8 Å and ~4.0 Å. An initial 6 Å map was obtained from phases from a Ta_6Br_{12} derivative; these phases helped locate heavy atoms in other derivative data sets. Multiple isomorphous replacement with anomalous scattering (MIRAS) phases produced an excellent map to 4.1 Å where all three components were visible (Fig. 1, Supplementary Figs 2 and 3, and Supplementary Tables 2–4). Density was particularly clear for the RNA molecules, whereas density was only clear for the protein backbone and hence the high resolution model of the *T. maritima* protein[21] was positioned without significant rebuilding. The P RNA was built into the map using the structures of *T. maritima*[17] and *Bacillus stearothermophilus*[14] P RNA as guides, whereas *T. maritima* tRNA^{Phe} used yeast tRNA^{Phe} as a guide[22]. The structure was refined using anisotropic data to 3.8 Å resolution. Crystals with a tRNA leader present were obtained by soaking a short oligonucleotide with and without samarium chloride and this structure was refined to 4.2 Å resolution.

Overall structure

In the complex, the tRNA sits with the acceptor stem against RNase P, making several tRNA–P RNA intermolecular contacts (Fig. 1 and Supplementary Fig. 1). The TΨC and D loops of the tRNA contact the S-domain, while the acceptor stem extends from the S-domain into the C-domain crossing the main P1/P4/P5 coaxial stem (Fig. 2 and Supplementary Fig. 4). The 3′ CCA end of the tRNA enters a tunnel formed by P6/P15/P16/P17 and base pairs with nucleotides in the L15 region (Fig. 2 and Supplementary Figs 4 and 5), an interaction recognized previously[23]. The 5′ end of the tRNA indicates the location of the active site, which is close to the region where P4, P5 and CR-IV intersect. The protein component is also adjacent to the 5′ end of the

[1]Department of Molecular Biosciences, Northwestern University, Evanston, Illinois 60208, USA. [2]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637, USA. †Present addresses: Departamento de Bioquímica y Biología Estructural, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Ciudad Universitaria, Apartado Postal 70-243, México 04510, México (A.T.-L.); Abbott Laboratories, Abbott Park, Illinois 60064-6400, USA (K.K.S.).

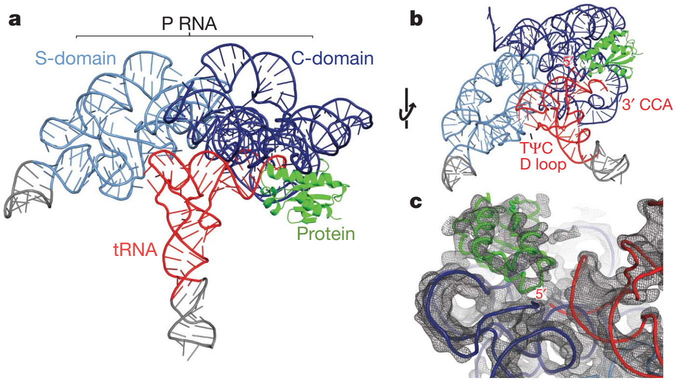


Figure 1 | Crystal structure of the *T. maritima* RNase P holoenzyme in complex with tRNA. **a**, Structure of bacterial RNase P, composed of a large RNA subunit (338 nucleotides, ~110 kDa) and a small protein component (117 amino acids, ~14.3 kDa), in complex with tRNA (76 nucleotides, ~26 kDa). The RNA component serves as the primary biocatalyst in the reaction and contains two domains, termed the catalytic (C, blue) and specificity (S, light blue) domains. The RNase P protein (green) binds the 5′ leader region of the pre-tRNA substrate and assists in product release. Transfer RNA ($tRNA^{Phe}$) (red) makes multiple interactions with the P RNA (see Fig. 2 and Supplementary Fig. 1 for details). Regions in grey denote additional RNA nucleotides required for crystallization. **b**, Alternative view of the RNase P–tRNA complex, identifying the tRNA recognition regions: the 5′ end where catalysis occurs, the 3′ CCA end, and the highly conserved TΨC and D loop regions. **c**, View of the 4.1 Å experimental electron density map centred on the 5′ end of tRNA. The map is represented as a dark grey mesh, contoured at 1.4 r.m.s.d.

tRNA, but does not contact it. The protein contacts include the CR-IV and CR-V regions, the P15 stem, and the P2/P3 helix interface (Fig. 3 and Supplementary Figs 1 and 6). The pre-tRNA leader makes extensive contacts with the protein, but few with the P RNA.

The components of the RNase P holoenzyme are largely unchanged when bound to tRNA (Supplementary Fig. 7). A comparison between *T. maritima* P RNA alone[17] and in the complex reveals an overall similar fold (backbone normalized root mean square deviation (r.m.s.d.) ~1.1 Å) with a small change in the relative orientation of the two domains (Supplementary Fig. 7). The only major change in the P RNA structure occurs in the vicinity of the P15–P17 stems (Supplementary Figs 8 and 9). A few additional residues at the amino terminus were clear and follow a similar path to the *B. subtilis* protein[24] (Supplementary Fig. 10); no changes in the structure of the protein component were detected. The structures of yeast and *T. maritima* $tRNA^{Phe}$ show remarkable resemblance (backbone normalized r.m.s.d. for acceptor stem ~0.8 Å) (Supplementary Fig. 11). Further, a comparison with previous models reveals an excellent agreement with the predicted secondary structure[25] and a good agreement with the models of the complex[19,26,27] (Supplementary Fig. 12).

tRNA recognition

The observed RNA–RNA interactions involved in substrate recognition agree with previous biochemical studies[5,23,28] and include (1) stacking between bases in the tRNA TΨC and D loops and the P RNA S-domain, (2) an A-minor interaction at the acceptor stem, and (3) the formation of canonical base pairs at the 3′ end of tRNA (Fig. 2 and Supplementary Fig. 1). The first interaction identifies the TΨC loop as a key element in recognition. Both the tRNA D and TΨC loops have unstacked bases (G19 and C56) that interact with unstacked bases in

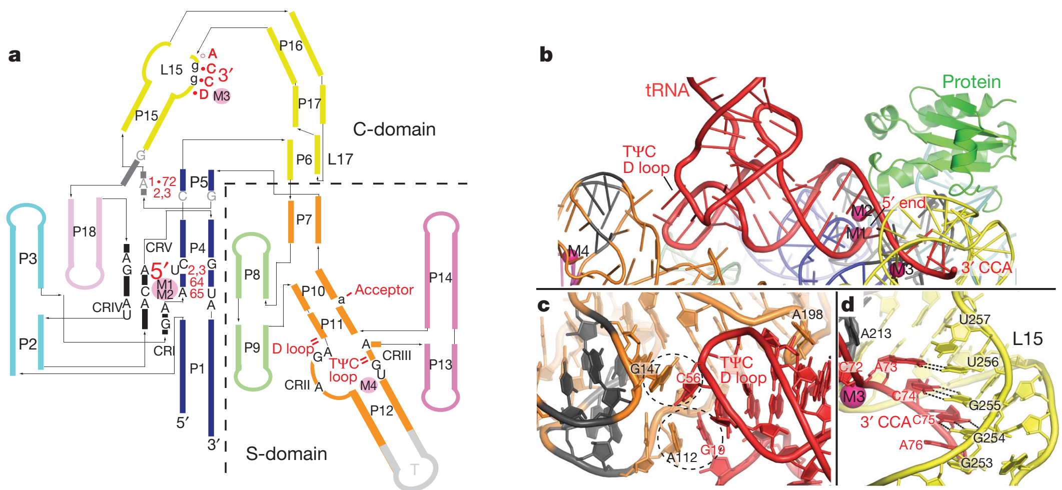


Figure 2 | tRNA recognition by RNase P is mediated by RNA–RNA interactions. **a**, Schematic of the P RNA secondary structure mapping the tRNA–P RNA contacts observed in the crystal structure. The tRNA nucleotides (1•72, 2, 3, 64 and 65) and regions (5′, 3′, TΨC loop, D loop and acceptor) involved in direct interactions are shown in red. Intermolecular base pairs form between the 3′ end of tRNA (DCCA) and loop 15 (L15), where D is the discriminator nucleotide that serves as an identity element in tRNA biogenesis. P RNA nucleotides that are universally conserved (black, uppercase), conserved among all bacteria (grey, uppercase), or highly conserved in bacteria (black, lowercase) are identified. Metal ions are shown as filled pink circles, and denote the location of the active site (M1, M2), and other structurally important regions (M3, M4). Single and double dashes in red represent minor groove and base stacking interactions, respectively. All identified tRNA–P RNA contacts are within 4 Å. The crystallized *T. maritima* P RNA consists of eighteen paired helices (P), five universally conserved regions (CR-I to CR-V) (black), two junctions containing conserved nucleotides in bacteria (dark grey), several loop (L) regions, and an engineered tetraloop region (T, light grey). The coaxial P1/P4/P5 stem is shown in blue, P2/P3 stems in cyan, P6/P15/P16 and L15/L17 in yellow, P7 and P10/P11/P12 in orange, P8/P9 in light green, and P13/P14 in pink (see Supplementary Fig. 1 for additional details). **b**, Recognition of tRNA by the P RNA of RNase P. The acceptor stem of tRNA (red) docks onto the P RNA (coloured as in **a**) making a series of interactions, including base stacking in the TΨC/D loops of tRNA and the S-domain, an A-minor interaction, and base pairing, ribose zipper and stacking interactions between the 5′ and 3′ ends of tRNA and the C-domain. The protein (green) makes no direct contacts with mature tRNA. Critical metal ions (M1–M4) identified are shown as magenta spheres. **c**, tRNA recognition by the S-domain. Two universally conserved P RNA regions (CR-II and III, dark grey) facilitate base stacking interactions with unstacked bases in the structurally conserved TΨC and D loops of tRNA. Dashed circles highlight this stacking interaction between P RNA residues A112, G147 and tRNA residues G19, C56. A conserved P RNA adenosine (A198) stacks into the minor groove of the acceptor tRNA stem. **d**, Recognition of the tRNA 3′ CCA by the C-domain. Intermolecular base pairs form between the 3′ tRNA (ACC) and the L15 (GGU) loop of P RNA. This interaction is stabilized by a structural metal (M3, magenta sphere) and a L15 ribose zipper conformation.

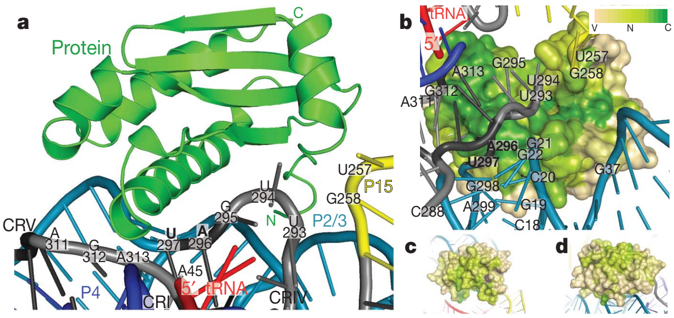


Figure 3 | Protein–RNA contacts within the RNase P holoenzyme. **a**, The protein sits on the P RNA surface formed by conserved regions I, IV and V. The protein (green, shown as ribbons) additionally contacts the L15/P15 junction and the P2/3 helices (P RNA as coloured in Fig. 2). Labelled P RNA nucleotides make protein contacts (within 4 Å) and include: A45 in CR-I, U257 and G258 in the L15/P15 junction, U293, U294, G295, A296 and U297 in CR-IV, and A311, G312 and A313 in CR-V. Bold nucleotides are universally conserved. **b**, Surface representation of the protein coloured by sequence conservation (variable (V), tan; neutral (N), light green; conserved (C), green). A highly conserved patch in the protein extends from the vicinity of the 5′ end of the tRNA, and interacts with P RNA conserved regions IV (U293-U297) and V (A311-A313). Other P RNA nucleotides that make protein contacts include: the P2 helix (C18-G22, G298-A299), the P3 helix (G37) and the L15/P15 junction (U257-G258). Four hundred and ninety bacterial RNase P proteins were included in the analysis of the sequence conservation using the ConSurf server[48]. Panels **c** and **d** show different orientations to emphasize that high sequence conservation is concentrated in the region of the protein that faces the conserved regions of the P RNA. Neutral or slightly conserved regions shown in these two orientations correspond to a patch that interacts with the leader.

the P RNA (A112 and G147), forming G19-A112 and C56-G147 stacks in the complex. The second major interaction involves a highly conserved unstacked adenosine (A198) in the P11 stem entering the minor groove of the tRNA acceptor stem. These interactions facilitate shape complementarity and help explain the central role of the S-domain in recognition. The third major interaction involves intermolecular base pairing between the tRNA 3′ DCCA motif and the L15 loop. This interaction is probably conserved in all bacterial and most archaeal RNase Ps, but not in organisms where CCA is added post-transcriptionally[1]. The fourth to last nucleotide, A73, forms a Watson–Crick base pair with nucleotide U256. C74 and C75 form Watson–Crick base pairs with G255 and G254, while the terminal A76 forms a weak interaction with G253. To accommodate these intermolecular base pairs, the two strands of L15 fold into a ribose zipper. In addition, a structural metal ion (M3) (Fig. 2 and Supplementary Fig. 13) binds adjacent to this P RNA-tRNA region and is likely to correspond to a metal ion identified biochemically[7]. In the complex, the 3′ end of the tRNA separates from the 5′ end and enters a wide opening formed by P6/P15/P16/P17 (Figs 1 and 2 and Supplementary Figs 4 and 5). This opening is ~20 Å in diameter, can easily accommodate a single-stranded RNA molecule, and is created when the P6/L17 pseudoknot forms (Supplementary Figs 1 and 5).

Protein–RNA interactions

The bacterial RNase P protein structure is highly conserved, but has little or no sequence or structural similarity with the protein components of archaea or eukarya[29]. In the complex, the protein is near the 5′ end of tRNA, but is too far (over 6 Å) to make direct contacts. The protein sits between the P15 and P3 stems (Fig. 3 and Supplementary Fig. 6), and also contacts the CR-IV and CR-V loop regions of P RNA. Comparison of bacterial sequences shows that the protein has a large, contiguous area with high sequence conservation (Fig. 3 and Supplementary Fig. 6) including important residues identified previously[11,30,31]. The conserved area extends in an arch along the surface of the protein, starting from a point close to the 5′ end of the tRNA and faces the universally conserved modules.

To investigate the interactions with the leader, crystals were soaked with a short oligoribonucleotide in the presence and absence of Sm^{3+}. Fourier difference maps to 4.2 Å show five phosphates of the leader along the conserved surface of the protein (Fig. 4 and Supplementary Fig. 6), but the position of the nucleobases was ambiguous. The structure shows that the leader contacts residues Phe 17, Phe 21, Lys 51, Arg 52 and Lys 90 and probably interacts with Gln 28, Lys 56 and Arg 89, in agreement with biochemical results[6,9,12,30,31]. The 3′ end of the leader is located adjacent to the 5′ end of the tRNA and near two conserved residues (Arg 52 and Lys 56). A metal ion is present in between the leader 3′ and the 5′ end of mature tRNA (Figs 4 and 5 and Supplementary Fig. 14), but is too distant (>4 Å) to ligate protein residues directly. Leader nucleotides −1 to −3 are poised to interact with nucleotides A213, U294, G295 and A314 of P RNA (Fig. 4 and Supplementary Figs 1 and 6). These results indicate that the major role of the protein component is to interact with the leader to align the pre-tRNA in the complex, as observed previously[9,11,31].

Active site

The location of the active site is inferred from the 5′ end of mature tRNA (Fig. 5 and Supplementary Fig. 15). The phosphate backbone of tRNA nucleotides (+1 to +3) sits on the major groove of the P4 stem (near A50, G51 and U52), and places the tRNA 5′ end next to the P4

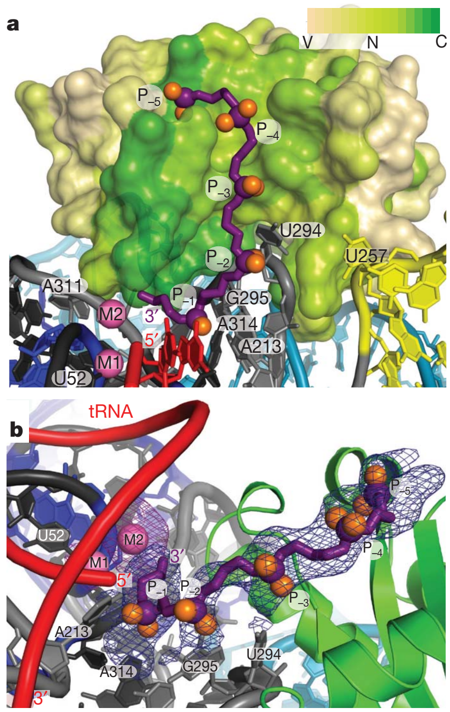


Figure 4 | Pre-tRNA leader–protein interactions in the RNase P holoenzyme. **a**, Surface representation of the protein coloured by sequence conservation as in Fig. 3. The pre-tRNA 5′ leader (purple, with purple and orange spheres for the phosphorous and non-bridging oxygens, respectively) was modelled as a polyphosphate chain with five phosphates (P_{-1} to P_{-5}). The leader follows a highly conserved patch in the protein extending from the 5′ end of the mature tRNA (red) and away from the P RNA. The addition of a 5′ leader with metal (Sm^{3+}) reveals a second metal ion (M2). **b**, Alternative view of the pre-tRNA leader–protein interaction. Each phosphate position (P_{-1} though P_{-5}) was visible in a 4.2 Å difference Fourier map (mF_o − DF_c) calculated from crystals where only the leader was soaked into the crystals (blue mesh, 3 r.m.s.d. contour levels). A second 4.2 Å difference Fourier map (mF_o − DF_c) calculated from crystals where the leader and Sm^{3+} metal were soaked into the crystals shows clearly the position of the second metal ion (magenta mesh, 3.5 r.m.s.d. contour level). P RNA residues poised to make contacts are labelled. Nucleotide U52 serves as a reference point in **a** and **b** and does not interact with the 5′ leader oligonucleotide.

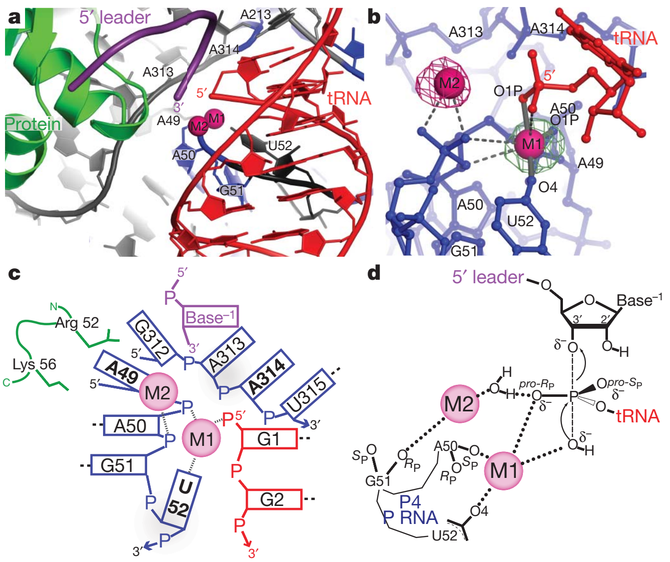


Figure 5 | Structure of the RNase P active site environment. **a**, The active site is inferred from the location of the mature 5′ end of tRNA. The diagram shows the position of the mature tRNA (red), the leader (purple), the protein component (green) and the P RNA (blue and grey). A group of conserved P RNA nucleotides (A49–U52, A213, A313 and A314) form part of the active site. Two metal ions (magenta spheres) are found in the active site. **b**, The two active site metal ions (M1 and M2) are within 4 Å of the 5′ phosphate of tRNA and the M1–M2 metal–metal distance is ~4.8 Å. The M1 metal makes contacts (≤2.1 Å, solid grey bonds, labelled) with tRNA (G1 O1P) and P RNA (A50 O1P and U52 O4) oxygens. Other possible ligands within 3.5 Å of M1 or M2 are represented by dashed grey lines (Supplementary Table 5). The figure shows two isomorphous difference Fourier ($mF_o - DF_c$) maps. The green mesh corresponds to a Eu^{3+} soak in the absence of leader and is contoured at the 9.5 r.m.s.d. level. The magenta mesh corresponds to a Sm^{3+} and 5′ leader soak and is contoured at the 5.5 r.m.s.d. level. The second metal is clearly visible only when the leader is present. **c**, Schematic diagram of the interactions around the active site. The diagram shows all residues within 8 Å of the 5′ phosphorus atom of tRNA. Short dashed lines represent metal ligand distances within 2.2 Å and longer dashed lines represent nucleotides which form canonical base pairs. Nucleotides in bold are universally conserved in P RNA. The P RNA, tRNA, 5′ leader, and protein side chains are shown in blue, red, purple and green, respectively. **d**, Proposed reaction mechanism for the endonucleolytic cleavage of pre-tRNA by RNase P based on the structure of the enzyme–product (E–P) complex and previous mechanistic studies[49,50]. The M1 metal distance to the 5′ phosphate ligands (Supplementary Table 5) in the E–P complex is consistent with the proposed enzyme–substrate (E–S) transition state. In this proposed reaction scheme, M1 is ~180° from the apical O3′ position and activates a hydroxyl nucleophile for an in-line nucleophilic displacement, creating a new bond and displacing the 3′ scissile phosphate oxygen. As RNase P proceeds through an S_N2 reaction pathway, the stereochemistry around the phosphorus atom undergoes a net inversion of configuration. If the *pro-R_P* (O2P) oxygen coordinates metal in the E–S complex during catalysis, as previously observed[49,50], this would subsequently allow for the *pro-S_P* (O1P) oxygen to coordinate metal in the E–P complex, as observed in the crystal structure. Product release could be facilitated by a metal (M2) coordinated water, which would enable proton transfer to the 3′ scissile oxygen. The exact active site geometry and identity of other metal ligands in an E–S complex has yet to be established.

phosphate backbone and nucleotides A313 and A314 (Fig. 5 and Supplementary Fig. 15). The universally conserved U52 nucleotide is unstacked from the P4 stem and faces the tRNA 5′ end. In addition, the tRNA 1•72 base pair is stabilized by an adenosine stack with A213, a nucleotide conserved in all bacteria.

A metal ion (M1), putatively magnesium, is found trapped between the tRNA 5′ end, the A50 and G51 phosphates, and the O4 oxygen of the universal U52 nucleotide and was confirmed using crystals soaked with Sm^{3+} and Eu^{3+} (Supplementary Figs 14 and 15). Putative M1 metal contacts include the A50 non-bridging phosphoryl oxygen, the O4 oxygen of the U52 nucleobase, and the O1P oxygen at the 5′ end of tRNA. Other metal-ligand interactions may include: the backbone of A50, the phosphoryl oxygen of G51, and the 5′ end of tRNA (Supplementary Table 5). Many of these oxygen ligands have been implicated in metal coordination and catalysis[32–34]. The M1 site may also coincide with a site (M6) observed in the structure of *B. stearothermophilus* P RNA[35]. The structure of the complex suggests that M1 participates in catalysis by directly binding P RNA and the 5′ phosphate of tRNA.

A second metal (M2) was located in experiments where the leader was soaked in the presence of Sm^{3+}. The M2 metal is in close proximity to the phosphoryl oxygens of G51, the O3′ of the leader, and the 5′ end of tRNA (Supplementary Table 5). The two metals observed in crystals soaked with the leader and Sm^{3+} are ~4.8 Å apart (Fig. 5 and Supplementary Fig. 15). The structures indicate that the active site includes at least two metal ions upon complex formation with pre-tRNA. Due to its location, the M2 metal ion could make additional contacts with both the tRNA and the P RNA during catalysis.

The structures of the active site of the complex and the apo-ribozyme structures are similar (Supplementary Figs 7, 16 and 17), including the presence of a metal ion next to the P4 helix[35]. With the exception of the U52 nucleobase (Supplementary Fig. 16), no large changes are observed in the active site region. A fully occupied M2 site is observed only in the presence of leader, suggesting that a local metal-dependent conformation change may occur, as previously reported[6]. The structure also reveals that the tRNA 5′ and 3′ ends splay and separate to interact with the P RNA (Supplementary Fig. 11), confirming the need for movement of the tRNA ends[36,37]. Although accommodating the upstream RNA leader probably requires local protein and P RNA structural changes, the location of the active site is not significantly altered and is largely pre-assembled.

Mechanistic implications

RNase P can cleave a variety of substrates[1,10,38], but pre-tRNA is the only one that is common among all organisms. To decipher its function, it is important to understand two different aspects of pre-tRNA processing by RNase P: substrate specificity and the chemical mechanism of cleavage.

tRNA recognition by RNase P involves the highly conserved tRNA TΨC and D loops and the CR-II and CR-III in the S-domain of P RNA. Thus, regions with high sequence and structure conservation are involved in specific tertiary interactions, suggesting a universal mode of recognition among all RNase P. The presence of unpaired nucleotides next to the cleavage site is also an important feature for pre-tRNA recognition, although it is unclear whether this is a universal feature of all natural substrates[1]. Finally, pre-tRNA is usually processed to form a 7-base-pair-long acceptor stem. An additional role of the interactions between CR-II and CR-III and tRNA may be to serve as a 'ruler' that ensures that the correct lengths are processed, although there is some flexibility as tRNAs with acceptor stems 8 base pairs long can be processed[39]. The interaction with the 3′ CCA end is also a key recognition feature, but may not be necessarily an RNA–RNA interaction in higher organisms. The L15 loop of P RNA is not found in eukarya or some archaea[40] and its function may be replaced by additional protein(s), suggesting that 3′ CCA intermolecular base pairing is not a universal interaction.

The second important aspect of RNase P function is the chemical mechanism of cleavage. Hydrolysis of a phosphodiester bond generates the mature 5′ RNA product. Whereas it is not possible to propose a complete mechanism from a structure at this resolution, the RNase P–tRNA structures, together with extensive biochemical information, help identify the major active site components. The structure indicates that at least two distinct metals play a direct role. It is possible to propose a transition state model (Fig. 5d) where the M1 metal directly positions the scissile phosphate oxygens of the substrate and enables a hydroxyl ion to perform an S_N2-type nucleophilic substitution. In this scenario,

the M2 metal ion stabilizes the transition state and mediates proton transfer to the 3′ scissile oxygen during product release, as proposed previously[7]. Other universally conserved nucleotides in the vicinity seem to have a structural role in forming the correct structure and are not directly involved in catalysis, consistent with proposals that sequence conservation is largely the result of strong structural constraints[19]. Hence, the RNase P–tRNA complex shows how the P RNA structure can serve as a scaffold to bind and orient metals and substrate properly. It seems that RNase P uses a two-metal ion catalytic mechanism, similar to other mechanisms proposed based on other large ribozyme structures[41,42] and originally put forth as a general mechanism for many ribozymes[43].

The structural studies of the holoenzyme–tRNA complex help to show that all RNase P ribozymes share a common, RNA-based mechanism of RNA cleavage and recognition that involves two universally conserved structural modules. Adaptation through the addition of protein increases RNase P functionality by positioning accurately the 5′ leader pre-tRNA substrate and by contacting conserved regions of the P RNA structure. The unique tertiary fold of the P RNA uses shape complementarity, specific RNA–RNA contacts, and intermolecular base pairing to recognize its substrate efficiently. Within this tertiary fold, the universally conserved regions are crucial to form the active site scaffold and to create regions involved in tRNA recognition. In addition, both P RNA and the pre-tRNA help to coordinate two catalytically important metal ions essential for the putative mechanism of pre-tRNA cleavage. The RNase P–tRNA complex offers a glimpse into the transition from an ancient, RNA-based world to the present, protein-catalyst dominated world and affirms that RNA molecules can display comparable versatility and complexity.

METHODS SUMMARY

Crystallization. Preparation, purification and folding of *T. maritima* RNase P and $tRNA^{Phe}$ have been described[8,17,44]. For crystallization, the components were mixed in a 1:1.1:1 (P RNA:pre-tRNA:protein) molar ratio to a concentration of 45 μM. The mixture was heated to 94 °C (2 min), cooled to 4 °C (2 min), and after the addition of $MgCl_2$ to a final 10 mM concentration, further incubated at 50 °C (10 min) and 37 °C (40 min). Crystals were obtained by mixing 1 μl of complex with 1 μl of reservoir solution (1.8 M Li_2SO_4, 50 mM sodium cacodylate (pH 6.0)) and equilibrated by vapour diffusion at 30 °C. Crystals were cryo-protected using reservoir solution containing 15% xylitol.

Data collection and structure determination. Diffraction data were collected at 100 K at the LS-CAT sector at the APS. Complete native and Ta_6Br_{12}, $SmCl_3$, $EuCl_3$ and iridium hexammine $(Ir(NH_3)_6)^{3+}$ derivatives were collected. A weak Molecular Replacement[45] solution using a trimmed model of the tRNA–P RNA complex[19] located the Ta_6Br_{12} cluster. Multi-wavelength anomalous dispersion (MAD) phases[46] from the cluster extended to ~6 Å, with the map showing a clear envelope. These phases were used to locate the other heavy atoms that were used to calculate a 4.1 Å MIRAS map. To locate the pre-tRNA leader, crystals were soaked with a *T. maritima* 5′ tRNA 7-nucleotide leader sequence (final concentration 0.2 mM), with and without 14 mM $SmCl_3$. Difference maps allowed the placement of five pre-tRNA nucleotides and the unambiguous identification of a second active site metal. The experimental electron density map was of excellent quality and allowed model building of nearly all RNA phosphate and nucleobase positions and accurate placing of the protein. Model building was guided by the known structures[14,17,21]. Final R_{work} and R_{free} are 24.9% and 27.0%, respectively, with r.m.s.d. of 0.007 Å and 1.24° for bonds and angles. Figures were made with PyMOL[47].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 June; accepted 17 September 2010.

Published online 14 November 2010.

1. Hartmann, R. K., Gossringer, M., Spath, B., Fischer, S. & Marchfelder, A. The making of tRNAs and more – RNase P and tRNase Z. *Prog. Mol. Biol. Transl. Sci.* **85,** 319–368 (2009).
2. Kazantsev, A. V. & Pace, N. R. Bacterial RNase P: a new view of an ancient enzyme. *Nature Rev. Microbiol.* **4,** 729–740 (2006).
3. Liu, F. & Altman, S. *Protein Reviews Volume 10: Ribonuclease P* (Springer, 2010).
4. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35,** 849–857 (1983).
5. Pan, T., Loria, A. & Zhong, K. Probing of tertiary interactions in RNA: 2′-hydroxyl-base contacts between the RNase P RNA and pre-tRNA. *Proc. Natl Acad. Sci. USA* **92,** 12510–12514 (1995).
6. Hsieh, J. *et al.* A divalent cation stabilizes the active conformation of the B. subtilis RNase P pre-tRNA complex: a role for an inner-sphere metal ion in RNase P. *J. Mol. Biol.* **400,** 38–51 (2010).
7. Kirsebom, L. A. in *Protein Reviews Volume 10: RNase P* (eds Liu, F. & Altman, S.) Ch. 7, 113–134 (Springer, 2010).
8. Buck, A. H., Dalby, A. B., Poole, A. W., Kazantsev, A. V. & Pace, N. R. Protein activation of a ribozyme: the role of bacterial RNase P protein. *EMBO J.* **24,** 3360–3368 (2005).
9. Kurz, J. C., Niranjanakumari, S. & Fierke, C. A. Protein component of *Bacillus subtilis* RNase P specifically enhances the affinity for precursor-$tRNA^{Asp}$. *Biochemistry* **37,** 2393–2400 (1998).
10. Peck-Miller, K. A. & Altman, S. Kinetics of the processing of the precursor to 4.5 S RNA, a naturally occurring substrate for RNase P from *Escherichia coli*. *J. Mol. Biol.* **221,** 1–5 (1991).
11. Koutmou, K. S. *et al.* Protein-precursor tRNA contact leads to sequence-specific recognition of 5′ leaders by bacterial ribonuclease P. *J. Mol. Biol.* **396,** 195–208 (2010).
12. Sun, L., Campbell, F. E., Zahler, N. H. & Harris, M. E. Evidence that substrate-specific effects of C5 protein lead to uniformity in binding and catalysis by RNase P. *EMBO J.* **25,** 3998–4007 (2006).
13. Reich, C., Olsen, G. J., Pace, B. & Pace, N. R. Role of the protein moiety of ribonuclease P, a ribonucleoprotein enzyme. *Science* **239,** 178–181 (1988).
14. Kazantsev, A. V. *et al.* Crystal structure of a bacterial ribonuclease P RNA. *Proc. Natl Acad. Sci. USA* **102,** 13392–13397 (2005).
15. Krasilnikov, A. S., Xiao, Y., Pan, T. & Mondragón, A. Basis for structural diversity in homologous RNAs. *Science* **306,** 104–107 (2004).
16. Krasilnikov, A. S., Yang, X., Pan, T. & Mondragón, A. Crystal structure of the specificity domain of ribonuclease P. *Nature* **421,** 760–764 (2003).
17. Torres-Larios, A., Swinger, K. K., Krasilnikov, A. S., Pan, T. & Mondragón, A. Crystal structure of the RNA component of bacterial ribonuclease P. *Nature* **437,** 584–587 (2005).
18. Chen, J.-L. & Pace, N. R. Identification of the universally conserved core of ribonuclease P RNA. *RNA* **3,** 557–560 (1997).
19. Torres-Larios, A., Swinger, K. K., Pan, T. & Mondragón, A. Structure of ribonuclease P–a universal ribozyme. *Curr. Opin. Struct. Biol.* **16,** 327–335 (2006).
20. Ferré-d'Amaré, A. R., Zhou, K. & Doudna, J. A. A general module for RNA crystallization. *J. Mol. Biol.* **279,** 621–631 (1998).
21. Kazantsev, A. V. *et al.* High-resolution structure of RNase P protein from *Thermotoga maritima*. *Proc. Natl Acad. Sci. USA* **100,** 7497–7502 (2003).
22. Shi, H. & Moore, P. B. The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *RNA* **6,** 1091–1105 (2000).
23. Kirsebom, L. A. & Svard, S. G. Base pairing between *Escherichia coli* RNase P RNA and its substrate. *EMBO J.* **13,** 4870–4876 (1994).
24. Stams, T., Niranjanakumari, S., Fierke, C. A. & Christianson, D. W. Ribonuclease P protein structure: evolutionary origins in the translational apparatus. *Science* **280,** 752–755 (1998).
25. Massire, C., Jaeger, L. & Westhof, E. Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. *J. Mol. Biol.* **279,** 773–793 (1998).
26. Buck, A. H., Kazantsev, A. V., Dalby, A. B. & Pace, N. R. Structural perspective on the activation of RNase P RNA by protein. *Nature Struct. Mol. Biol.* **12,** 958–964 (2005).
27. Tsai, H. Y., Masquida, B., Biswas, R., Westhof, E. & Gopalan, V. Molecular modeling of the three-dimensional structure of the bacterial RNase P holoenzyme. *J. Mol. Biol.* **325,** 661–675 (2003).
28. LaGrandeur, T. E., Huttenhofer, A., Noller, H. F. & Pace, N. R. Phylogenetic comparative chemical footprint analysis of the interaction between ribonuclease P RNA and tRNA. *EMBO J.* **13,** 3945–3952 (1994).
29. Mondragón, A. in *Protein Reviews Volume 10: Ribonuclease P* (eds Liu, F. & Altman, S.) Ch. 4, 63–78 (Springer, 2010).
30. Biswas, R., Ledman, D. W., Fox, R. O., Altman, S. & Gopalan, V. Mapping RNA-protein interactions in ribonuclease P from *Escherichia coli* using disulfide-linked EDTA-Fe. *J. Mol. Biol.* **296,** 19–31 (2000).
31. Niranjanakumari, S., Stams, T., Crary, S. M., Christianson, D. W. & Fierke, C. A. Protein component of the ribozyme ribonuclease P alters substrate recognition by directly contacting precursor tRNA. *Proc. Natl Acad. Sci. USA* **95,** 15212–15217 (1998).
32. Christian, E. L., Kaye, N. M. & Harris, M. E. Helix P4 is a divalent metal ion binding site in the conserved core of the ribonuclease P ribozyme. *RNA* **6,** 511–519 (2000).
33. Crary, S. M., Kurz, J. C. & Fierke, C. A. Specific phosphorothioate substitutions probe the active site of *Bacillus subtilis* ribonuclease P. *RNA* **8,** 933–947 (2002).
34. Christian, E. L., Smith, K. M., Perera, N. & Harris, M. E. The P4 metal binding site in RNase P RNA affects active site metal affinity through substrate positioning. *RNA* **12,** 1463–1467 (2006).
35. Kazantsev, A. V., Krivenko, A. A. & Pace, N. R. Mapping metal-binding sites in the catalytic domain of bacterial RNase P RNA. *RNA* **15,** 266–276 (2009).
36. Pomeranz Krummel, D. A., Kent, O., MacMillan, A. M. & Altman, S. Evidence for helical unwinding of an RNA substrate by the RNA enzyme RNase P: use of an interstrand disulfide crosslink in substrate. *J. Mol. Biol.* **295,** 1113–1118 (2000).
37. Gaur, R. K., Hanne, A., Conrad, F., Kahle, D. & Krupp, G. Differences in the interaction of *Escherichia coli* RNase P RNA with tRNAs containing a short or a long extra arm. *RNA* **2,** 674–681 (1996).
38. Forster, A. C. & Altman, S. Similar cage-shaped structures for the RNA components of all ribonuclease P and ribonuclease MRP enzymes. *Cell* **62,** 407–409 (1990).

39. Burkard, U. & Soll, D. The unusually long amino acid acceptor stem of *Escherichia coli* selenocysteine tRNA results from abnormal cleavage by RNase P. *Nucleic Acids Res.* **16,** 11617–11624 (1988).
40. Walker, S. C. & Engelke, D. R. Ribonuclease P: the evolution of an ancient RNA enzyme. *Crit. Rev. Biochem. Mol. Biol.* **41,** 77–102 (2006).
41. Stahley, M. R. & Strobel, S. A. Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* **309,** 1587–1590 (2005).
42. Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* **320,** 77–82 (2008).
43. Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl Acad. Sci. USA* **90,** 6498–6502 (1993).
44. Krivenko, A. A., Kazantsev, A. V., Adamidi, C., Harrington, D. J. & Pace, N. R. Expression, purification, crystallization and preliminary diffraction analysis of RNase P protein from *Thermotoga maritima*. *Acta Crystallogr. D* **58,** 1234–1236 (2002).
45. McCoy, A. J. *et al. Phaser* crystallographic software. *J. Appl. Cryst.* **40,** 658–674 (2007).
46. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276,** 472–494 (1997).
47. DeLano, W. L. The PyMOL molecular graphics system. 〈http://www.pymol.org〉 (2002).
48. Landau, M. *et al.* ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33,** W299–W302 (2005).
49. Chen, Y., Li, X. & Gegenheimer, P. Ribonuclease P catalysis requires Mg^{2+} coordinated to the *pro*-R_P oxygen of the scissile bond. *Biochemistry* **36,** 2425–2438 (1997).
50. Warnecke, J. M. *et al.* Ribonuclease P (RNase P) RNA is converted to a Cd^{2+}-ribozyme by a single Rp-phosphorothioate modification in the precursor tRNA at the RNase P cleavage site. *Proc. Natl Acad. Sci. USA* **93,** 8924–8928 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Sontheimer and O. Uhlenbeck for comments and suggestions, N. Pace for the gift of the *T. maritima* RNase P protein plasmid, Obiter Research, A. Davis and M. E. Duban for advice and preparation of iridium hexammine, and A. Samelson for discussions and assistance. In addition, we are grateful for data collection assistance from S. Anderson, Z. Wawrzak, and staff at LS-CAT. Research was supported by the NIH. N.J.R. is an NRSA postdoctoral fellow.

Author Contributions A.M. directed the work. A.T.-L. and A.M. conceived the project. All authors performed and designed experiments. N.J.R. obtained crystallographic data. N.J.R. and A.M. analysed the crystallographic data. N.J.R. and A.M. wrote the paper with contributions from all authors.

Author Information Coordinates for the RNase P holoenzyme–tRNA complex, with and without 5′ tRNA leader sequence, have been deposited into the RCSB Protein Data Bank (accession code 3OKB and 3OK7, respectively). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.M. (a-mondragon@northwestern.edu).

METHODS

Preparation of the *T. maritima* RNase P holoenzyme–tRNA^{Phe} ternary complex. RNA transcriptions were performed *in vitro* using purified His_6-tagged T7 RNA polymerase using standard protocols[51]. Sequences from the *T. maritima* RNase P RNA and tRNA^{Phe} genes were inserted into a pUC19 vector at FokI and BsmAI restriction sites, respectively, allowing for run-off transcription of the DNA plasmid after digestion with the appropriate restriction enzyme (NEB). Constructions of modified RNA molecules with either mutations or additions were performed using a QuikChange mutagenesis kit (Stratagene). RNA samples were purified by 6% denaturing polyacrylamide gel electrophoresis (PAGE), identified by ultraviolet absorbance, recovered by diffusion into 50 mM potassium acetate (pH 7) and 0.2 M potassium chloride, and precipitated with ethanol. tRNA was further purified by anion exchange (MonoQ (5/50 GL)) and gel filtration (HiPrep 26/60, Sephacryl S-200) chromatography (GE Health Sciences). Over-expression and purification of the RNase P protein from *T. maritima* was performed as described previously[44].

To form the RNase P holoenzyme–tRNA complex, unfolded P RNA, unfolded tRNA and P protein molecules were mixed at a 1:1.1:1 molar ratio in 66 mM HEPES, 33 mM Tris (pH 7.4), 0.1 mM EDTA (1× THE) and 100 mM CH_3COONH_4 (Ref. 8). The ternary mix, at a final concentration of 45 μM, was incubated at 94 °C for 2 min and then cooled to 4 °C over 2 min. After addition of $MgCl_2$ to a final 10 mM concentration, the reaction mixture was incubated at 50 °C for 10 min, followed by incubation at 37 °C for 40 min, and finally cooled to 4 °C over 30 s.

Rational design of an RNA tertiary module to build a crystal lattice. To promote formation of a crystal lattice, intermolecular interactions were facilitated by introducing a tertiary structure interaction module. Based on the *T. maritima* RNA sequence and a proposed model of the P RNA–tRNA complex[19], constructs were designed where a tetraloop was inserted into the P12 loop (L12) of the P RNA and a tetraloop-receptor into the anticodon stem of tRNA (Supplementary Fig. 1). These two RNA regions were chosen as they were deemed to be far from the active site or other regions involved in specific interactions. In addition, the P12 stem of P RNA has a highly variable helix length across all organisms, lacks sequence conservation, and is non-essential or absent in several organisms[40]. The P12 and the anticodon loop of tRNA are not known to form any functional contacts. The length of the anticodon and the P12 stems were systematically varied by single base pair insertions adjacent to the tetraloop and tetraloop receptor module, thus altering the position (~2.7 Å per base pair added) and orientation (~36° per base pair added) of the tetraloop receptor and the tetraloop. Forty two combinations of molecules were screened for crystallization conditions using a sparse matrix approach employing a set of crystallization conditions developed locally. A few combinations of RNA molecules produced crystals, with most of them diffracting poorly. The best crystals were obtained from a construct where the P12 and anti-codon stems were elongated by five and three base pairs respectively. Insertion of two G-U wobble pairs adjacent to the tetraloop-tetraloop receptor module further improved diffraction, and also created a binding site for an iridium hexammine cation.

Crystallization and data collection. Crystals were obtained by mixing 1 μl of complex with 1 μl of reservoir solution (1.8 M $LiSO_4$, 50 mM sodium cacodylate (pH 6.0)) and equilibrated by vapour diffusion hanging or sitting drops at 30 °C. Gel analysis of washed crystals show that all three components were present (data not shown). Attempts to crystallize the complex in the absence of protein yielded no crystals. Crystals suitable for data collection grew in approximately 3 weeks and were cryo-cooled in liquid nitrogen immediately after transfer to reservoir solution containing 15% xylitol. Crystals of the RNase P holoenzyme–tRNA ternary complex suitable for data collection grew to approximately ~80–300 μM per side/edge, and diffract anisotropically to 3.8 Å in the best direction and ~4.0 Å in other directions. Crystals belong to space group $P3_121$ ($a = b = 169.3$ Å, $c = 185$ Å) and contain one molecule per asymmetric unit.

A series of derivatized crystals were also prepared by soaking in heavy metal compounds. Derivatives were prepared by soaking the crystals in mother liquor plus the derivative and incubating for 2–24 h before transferring them to cryoprotectant with the derivative present and freezing them in liquid nitrogen. Successful derivatizations were obtained by soaking the crystals in the following compounds: 2 mM Ta_6Br_{12}, 14 mM samarium chloride (Sm^{3+}), 14 mM europium chloride (Eu^{3+}), and 15 mM iridium hexammine $(Ir(NH_3)_6)^{3+}$. However, several of the compounds partially precipitated upon addition to the mother liquor solution and hence the final concentration is not known precisely. In addition, crystals with a leader present were obtained by soaking in a 0.2 mM heptamer oligonucleotide (5′-$A_{-7}A_{-6}G_{-5}G_{-4}C_{-3}G_{-2}U_{-1}$-3′) (Thermo Fisher) for 4 h with and without 14 mM samarium chloride present. The sequence was chosen by selecting the most common nucleotide in the *T. maritima* tRNA leaders at each position.

All diffraction data were collected at 100 K at the Life Science-Collaborative Access Team (LS-CAT) sector located at the Advance Photon Source (APS) using Rayonix CCD detectors. As the crystals are very radiation sensitive, the data collection range was optimized using the program MOSFLM[52] to collect the most complete native or anomalous data set using the minimal rotation range. Multi-wavelength anomalous dispersion (MAD) data were collected from a tantalum bromide cluster (Ta_6Br_{12}) derivative at three different wavelengths. Single or multiple wavelength anomalous dispersion data were also collected from the samarium chloride (Sm^{3+}), europium chloride (Eu^{3+}), and iridium hexammine $(Ir(NH_3)_6)^{3+}$ derivatives. Data were processed with XDS[53] and scaled with SCALA[54]. All other processing was done with programs from the CCP4 suite[54], except when noted. Data collection statistics for native and derivative data sets are shown in Supplementary Table 2.

In all cases, the diffraction limits of the data were anisotropic. The extent of the anisotropy was determined using the Anisotropy Server[55] and the data were treated in three different ways: (1) without any anisotropy correction; (2) carving the data to the limits suggested by the anisotropy server (3σ cut-off level on amplitudes); and (3) applying an anisotropic correction to the data using the server. For the second case, the integrated data from XDS was carved to the limits suggested by the server and then merged and scaled with SCALA before final processing. In many instances, the phasing and refinement calculations were done separately with the complete and carved data sets and the results compared. Overall, the different ways of treating the data had little effect on the final results, even though the data collection statistics were better for the carved data set (see Supplementary Tables 2 and 3).

Structure determination and model refinement. Molecular replacement (MR) studies with the program PHASER[45] using a proposed partial model of the P RNA–tRNA complex[19] gave a weak low resolution (25–8 Å) MR solution (*Z*-scores: 5.4 and 9.0 for the rotation and translation functions, respectively). Phases calculated from the MR solution were used to locate the position of the three sites in the Ta_6Br_{12} cluster data set. The program SHARP[46] was used to calculate MAD phases using data from three different wavelengths and spherically averaged form factors for the cluster. The solvent-flattened MAD map was of excellent quality but the phases were only good to ~6 Å resolution. The positions of the Eu^{3+}, Sm^{3+} and $(Ir(NH_3)_6)^{3+}$ heavy atoms were determined using the cluster phases. The parameters from the cluster and other derivatives could not be refined simultaneously and instead multiple isomorphous replacement with anomalous scattering (MIRAS) phases to 4.1 Å resolution were calculated using data from the single-atom derivatives together with phase information to 6 Å from the cluster data. The SOLOMON[56] solvent-flattened map was very clear (Supplementary Fig. 2) and all three molecules were apparent in the map. The model for the P RNA–tRNA complex[19] fit well in many areas, but the map showed regions where the model needed to be changed, regions that were missing in the model, like the P12 extension and the pseudoknot region, and the position of the protein. The models for the tRNA and P RNA were rebuilt completely using the high resolution model of yeast tRNA^{Phe} (ref. 22), *T. maritima* P RNA[17] and *B. stearothermophilus* P RNA[14] as guides. All regions of the RNA molecules were visible in the map and regions that were missing in the original *T. maritima* P RNA model were built. Some minor corrections to the original model were needed, but overall the models for P RNA agree well. The protein density was clear for the backbone, but not for the side chains and hence the high resolution model of the *T. maritima* protein[21] was placed on the experimental electron density map as a rigid body with minimal rebuilding.

Refinement was performed using Refmac5[57] and BUSTER[58]. Owing to the resolution of the data, the models were restrained to enforce good hydrogen bonding distance between Watson–Crick base pairs, planarity between base pairs (both for Watson–Crick and non-Watson–Crick base pairs), and C3′-endo sugar puckering for recognizable secondary structure elements. In addition, during BUSTER refinement the protein was restrained by the high resolution structure of the protein[21]. Model building with Coot[59] was interspersed with either Refmac5 or BUSTER refinement. During rebuilding, missing nucleotides were added as well as some missing residues at the N terminus of the protein. Mg^{2+} ions were included at positions that had high density peaks in residual maps and also coincided with heavy atom sites. Other large peaks in the native data set that coincided with phosphate positions in the leader-soaked crystals were modelled as phosphate ions. No individual atomic or group temperature factors were refined, only an overall anisotropic temperature factor. The final stages of the refinement were done with the program BUSTER. The refinement was done both with a carved data set where data outside the anisotropic diffraction limits (3σ cut-off) were excluded and also with a complete data set (isotropic) to the highest resolution limit (Supplementary Table 3). No significant difference was noted in the two refinements and the refinement statistics and electron density maps calculated from either data set were also virtually identical. It seems that the

anisotropic temperature factor correction in the refinement programs adequately modelled the modest anisotropy of the data.

The final model for the P RNA includes nucleotides 1 to 338. Only the phosphate backbone was modelled for nucleotides 39, 241 and 314–317. In addition, 9 nucleotides were inserted between nucleotides 130 and 136 to account for the extension added for crystallization (Supplementary Fig. 1). The final model for the tRNA includes nucleotides 1–76, but only the phosphate backbone was modelled for nucleotides 16, 17 and 20. The crystallization module added eight extra nucleotides incorporated at the end of the anticodon stem (Supplementary Fig. 1). Nearly the entire anticodon stem and anticodon loop were altered to accommodate the tetraloop receptor and altered anticodon loop. The protein model includes residues 6 to 117. The positions of all side chains were ambiguous in the map and were not rebuilt, but kept as much as possible as in the original 1.2 Å model (PDB ID 1NZ0) during refinement. Side chains that collided with the RNA were rebuilt when needed. There are four Mg^{2+} and two phosphate ions in the model of the complex. The final model to 3.8 Å resolution has an overall R_{work} of 24.9% and R_{free} of 27.0% with a root mean square deviation (r.m.s.d.) from target values of 0.007 Å and 1.24° for bonds and angles, respectively. The model in the presence of the leader includes an additional polyphosphate molecule with five phosphates and two Mg^{2+} ions coinciding with metals ions M1 and M2. A total of five Mg^{2+} ions were modelled into the complex that contains the 5′ polyphosphate leader backbone. The final model to 4.21 Å resolution has an overall R_{work} of 25.8% and R_{free} of 26.7% with an r.m.s.d. of 0.007 Å and 1.23° for bonds and angles, respectively (see Supplementary Tables 3 and 4).

Model superpositions were done with programs from the CCP4 suite[54], lsqman[60] and Coot[59]. Diagrams were made with PYMOL[47]. Coordinates and structure factors have been deposited in the PDB with accession numbers 3OK7 and 3OKB.

Activity assays of RNase P holoenzyme. Cleavage assays measuring k_{cat}/K_M under single turnover conditions were performed on the RNase P and pre-tRNA constructs that gave the best diffracting crystals. The pre-tRNA (with a single nucleotide leader (−1)) which yielded crystals and a control pre-tRNA (containing a *T. maritima* nine nucleotide leader (−9)) were radioactively labelled at their 5′ ends. Labelled substrates were purified over a 10% denaturing polyacrylamide gel and identified by ^{32}P-phosphorimaging. The holoenzyme was folded and cleavage reactions were performed in identical conditions as the folding reaction (1× THE, 10 mM $MgCl_2$, 0.1 M CH_3COONH_4, 37 °C). The enzyme activity of both the modified RNase P which gave crystals and the *T. maritima* wild-type RNase P were tested. The reaction was initiated by mixing pre-folded RNase P holoenzyme (25, 50 and 100 nM) with pre-folded pre-tRNA substrate (<4 nM), incubated for various times ($t = 0$, 0.25, 1, 4 and 16 min), and subsequently quenched by adding 9 M urea, 50 mM EDTA. All reaction mixtures were loaded directly on a 15% denaturing polyacrylamide gel which separated the substrate from the product(s). To observe unambiguously the products of the leader (−1) pre-tRNA, thin layer chromatography (TLC) was also performed with polyethyleneimine (PEI)-cellulose coated plates, where the quenched reaction mixture was spotted and run in a 5% acetic acid/100 mM NH_4Cl solution. The dried gels and the TLC plates were exposed to a phosphorimaging screen and the reaction profile was quantified by a phosphorimager (Fuji Medical) using ImageGauge software. A plot of the percentage of product over time gave the cleavage reaction rate for each concentration. Single turnover conditions assuming a first order reaction follow the equation $P = P_\infty(1 - e^{k_{obs}\,t})$, where P is the fraction of pre-tRNA cleaved, P_∞ is the fraction of uncleaved pre-tRNA at the end of the reaction, and k_{obs} is the observed reaction rate constant. By measuring k_{obs} at different concentrations it is possible to obtain K_{cat}/K_M assuming Michaelis–Menten kinetics.

51. Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15,** 8783–8798 (1987).
52. Leslie, A. G. W. in *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* **26,** 27–33 (1992).
53. Kabsch, W. Automatic indexing of rotation diffraction patterns. *J. Appl. Cryst.* **21,** 67–72 (1988).
54. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
55. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103,** 8060–8065 (2006).
56. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52,** 30–42 (1996).
57. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53,** 240–255 (1997).
58. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D* **60,** 2210–2221 (2004).
59. Emsley, P. & Cowtan, K. *Coot*: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
60. Kleywegt, G. J. Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr. D* **52,** 842–857 (1996).

ARTICLE

doi:10.1038/nature09472

Quantitative reactivity profiling predicts functional cysteines in proteomes

Eranthie Weerapana[1,2]*, Chu Wang[1,2]*, Gabriel M. Simon[1,2], Florian Richter[3,4], Sagar Khare[3,5], Myles B. D. Dillon[2], Daniel A. Bachovchin[1,2], Kerri Mowen[2], David Baker[3,4,5] & Benjamin F. Cravatt[1,2]

Cysteine is the most intrinsically nucleophilic amino acid in proteins, where its reactivity is tuned to perform diverse biochemical functions. The absence of a consensus sequence that defines functional cysteines in proteins has hindered their discovery and characterization. Here we describe a proteomics method to profile quantitatively the intrinsic reactivity of cysteine residues en masse directly in native biological systems. Hyper-reactivity was a rare feature among cysteines and it was found to specify a wide range of activities, including nucleophilic and reductive catalysis and sites of oxidative modification. Hyper-reactive cysteines were identified in several proteins of uncharacterized function, including a residue conserved across eukaryotic phylogeny that we show is required for yeast viability and is involved in iron-sulphur protein biogenesis. We also demonstrate that quantitative reactivity profiling can form the basis for screening and functional assignment of cysteines in computationally designed proteins, where it discriminated catalytically active from inactive cysteine hydrolase designs.

Large-scale scientific endeavours such as genome sequencing and structural genomics are providing a wealth of new information on the full complement of proteins present in eukaryotic and prokaryotic organisms. Many of these proteins, however, remain partly or completely unannotated with respect to their biochemical activities[1]. New methods are therefore needed to characterize protein function on a global scale. Much effort is currently devoted to the characterization of post-translational modification events because these covalent adducts can have profound and dynamic effects on protein activity[2]. Another frequently overlooked parameter that defines functional 'hotspots' in the proteome is amino acid side-chain reactivity, which can vary by several orders of magnitude for a given residue depending on local protein microenvironment. Methods to measure side-chain reactivity en masse directly in complex biological systems have not yet been described, and as such, the reactive landscape of the proteome remains largely unexplored.

Among the protein-coding amino acids, cysteine is unique owing to its intrinsically high nucleophilicity and sensitivity to oxidative modification. The pK_a of the free cysteine thiol is between 8 and 9, meaning that only slight perturbations in the local protein microenvironment can result in ionized thiolate groups with enhanced reactivity at physiological pH[3]. Diverse families of enzymes use cysteine-dependent chemical transformations, including proteases, oxidoreductases and acyltransferases[4]. In addition to its role in catalysis, cysteine is subject to several forms of oxidative post-translational modification, including sulphenation (SOH), sulphination (SO_2H), nitrosylation (SNO), disulphide formation and glutathionylation, which endow it with the ability to serve as a regulatory switch on proteins that is responsive to the cellular redox state[5].

Functional cysteines, regardless of whether they are catalytic residues or sites of post-translational modification, do not conform to a canonical sequence motif, which complicates their systematic identification and characterization. pK_a measurements can identify cysteine residues with heightened nucleophilicity (or 'hyper-reactive' cysteines[6,7]), but this requires purified protein and detailed kinetic and mutagenic experiments[7,8] that cannot be performed on a proteome-wide scale. Additional methods have been introduced to computationally predict redox-active cysteines[9], identify cysteines with specific modifications[10–14], and qualitatively inventory electrophile-modified cysteines in proteomes[15–18]. Some of these studies have provided suggestive evidence that nucleophilic cysteines may possess a variety of important functions[14–18], although the non-quantitative methods used in each case precluded a robust and systematic evaluation of this potential relationship. We adopted a different strategy to globally characterize cysteine functionality in proteomes based on quantitative reactivity profiling with isotopically labelled, small-molecule electrophiles.

Quantifying cysteine reactivity in proteomes

Our approach, termed isoTOP-ABPP (isotopic tandem orthogonal proteolysis–activity-based protein profiling), has four features to enable quantitative analysis of native cysteine reactivity (Fig. 1a): (1) an electrophilic iodoacetamide (IA) probe, to label cysteine residues in proteins, that also has (2) an alkyne handle for 'click chemistry' conjugation of probe-labelled proteins[19] to (3) an azide-functionalized TEV-protease recognition peptide containing a biotin group for streptavidin enrichment of probe-labelled proteins[20], and (4) an isotopically labelled valine for quantitative mass spectrometry (MS) measurements of IA-labelled peptides across multiple proteomes (Supplementary Fig. 1). After tandem on-bead proteolytic digestions with trypsin and TEV protease[15,20], probe-labelled peptides attached to isotopic tags are released and analysed by liquid-chromatography-high-resolution MS to identify IA-modified cysteines and quantify their extent of labelling based on MS2 and MS1 profiles, respectively. An isoTOP-ABPP ratio, R, is generated for each identified cysteine that reflects the difference in signal intensity between light and heavy tag-conjugated proteomes.

We first verified the accuracy of isoTOP-ABPP by labelling varying amounts of a mouse liver proteome (1×, 2×, 4×) with the IA probe

[1]The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037, USA. [2]Department of Chemical Physiology, The Scripps Research Institute, La Jolla, California 92037, USA. [3]Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. [4]Interdisciplinary Program in Biomolecular Structure and Design, University of Washington, Seattle, Washington 98195, USA. [5]Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA.
*These authors contributed equally to this work.

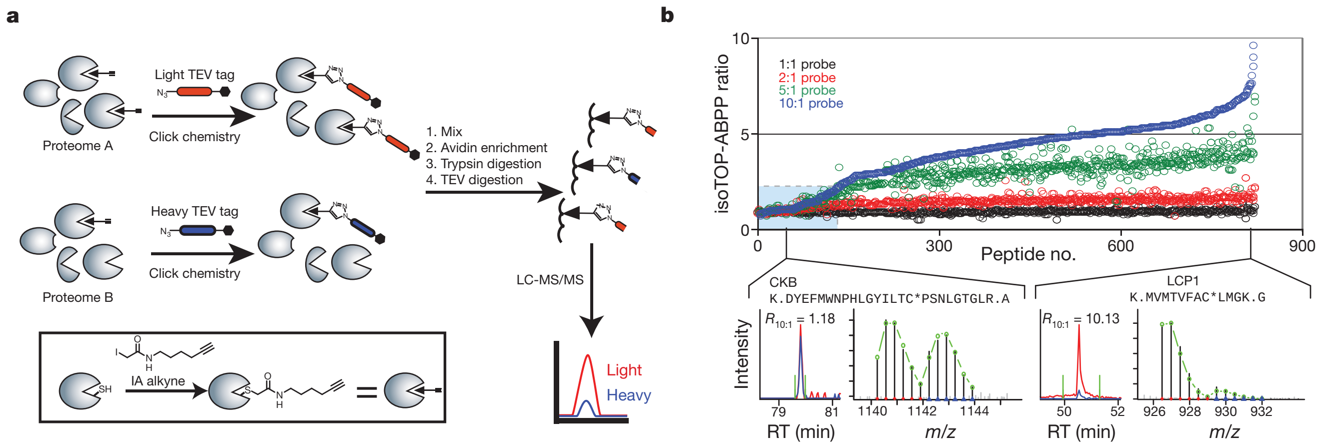


Figure 1 | A quantitative approach to globally profile cysteine reactivity in proteomes. a, isoTOP-ABPP involves proteome labelling, click-chemistry-based incorporation of isotopically labelled cleavable tags, and sequential on-bead protease digestions to provide probe-labelled peptides for MS analysis. The IA probe is shown in the inset. LC-MS/MS, liquid-chromatography-MS/MS. **b**, Measured isoTOP-ABPP ratios for peptides from MCF7 cells labelled with four pairwise IA probe concentrations (10:10 µM, 20:10 µM, 50:10 µM, 100:10 µM). The blue box highlights peptides with low isoTOP-ABPP ratios ($R < 2.0$). Chromatographs for creatine kinase B (CKB; low ratio) and plastin 2 (LCP1; high ratio) are shown, with elution profiles for heavy- and light-labelled peptides in blue and red, respectively, and green lines depicting peak boundaries used for quantification. Isotopic envelopes are shown for light- and heavy-labelled peptides with green lines representing predicted values. Sequences are shown for tryptic peptides containing IA-probe-labelled cysteines (marked by asterisks) in CKB and LCP1. RT, retention time. Additional chromatographs from isoTOP-ABPP experiments are in Supplementary Table 7.

followed by click chemistry conjugation with either the heavy or light variants of the azide-TEV-biotin tag. The observed signals for labelled cysteines closely matched the expected proteome ratios ($R_{1:1} \approx 1$, $R_{2:1} \approx 2$, or $R_{4:1} \approx 4$, respectively; Supplementary Fig. 2). A representative MS/MS profile of an IA-labelled peptide from our proteomic experiments is provided in Supplementary Fig. 3.

In contrast to traditional cysteine-alkylating protocols for proteomics that use millimolar concentrations of IA to stoichiometrically modify all cysteines in denatured proteins[21], we proposed that, by applying low (micromolar) concentrations of the IA probe to native proteomes, differences in the extent of alkylation would reflect differences in cysteine reactivity, rather than abundance. This hypothesis predicts that the reactivity of cysteines can be measured on a proteome-wide scale in isoTOP-ABPP experiments that compare low versus high concentrations of IA probe, where hyper-reactive cysteines would be expected to label to completion at low probe concentrations (generating isoTOP-ABPP ratios with $R_{[\mathrm{high}]:[\mathrm{low}]} \approx 1$) and less reactive cysteines should show concentration-dependent increases in IA-probe labelling (generating isoTOP-ABPP ratios with $R_{[\mathrm{high}]:[\mathrm{low}]} \gg 1$) (Supplementary Fig. 4). We tested this idea by performing four parallel isoTOP-ABPP experiments with the soluble proteome of the human breast cancer cell line MCF7 using pair-wise IA-probe concentrations of 10:10 µM, 20:10 µM, 50:10 µM and 100:10 µM (light:heavy). More than 800 probe-labelled cysteines were identified on 522 proteins, the vast majority of which exhibited escalating isoTOP-ABPP ratios (Fig. 1b) expected for reactions that did not reach completion over the tested probe concentration range. In contrast, a small subset of cysteines (<10%) showed nearly identical ratios at all probe concentrations tested ($R_{1:1} \approx R_{2:1} \approx R_{5:1} \approx R_{10:1} \approx 1$, Fig. 1b, shaded blue box). An expanded analysis of multiple human cancer line (Supplementary Fig. 5 and Supplementary Table 1) and mouse tissue (Supplementary Fig. 6 and Supplementary Table 2) proteomes treated with low (10 µM) and high (100 µM) IA-probe concentrations revealed consistent isoTOP-ABPP ratios for individual cysteine residues, indicating that the propensity of a cysteine to display high IA reactivity is an intrinsic property of the residue (and presumably its local protein environment), and not, in general, contingent on features specific to a particular cell or tissue. Additionally, isoTOP-ABPP ratios showed no correlation with either protein abundance or peptide ion intensity (Supplementary Fig. 7), indicating that they were independent of potential MS-based ionization sources for saturation. Finally, we confirmed that similar isoTOP-ABPP ratios were obtained for cysteines in reactions where time rather than the concentration of probe was varied (Supplementary Fig. 8 and Supplementary Table 3), confirming that low isoTOP-ABPP ratios reflect rapid reaction kinetics (hyper-reactivity), rather than saturable binding interactions (see Supplementary Discussion).

Hyper-reactivity predicts cysteine functionality

We next sought to assess the functional ramifications of the special subset of cysteines that showed hyper-reactivity in isoTOP-ABPP experiments. We first noted that multiple sites of IA-probe labelling on the same protein often showed markedly different isoTOP-ABPP ratios. For example, the glutathione *S*-transferase GSTO1 was labelled on four cysteine residues, three of which showed high ratios (C90, C192 and C237 had ratios of $R_{10:1} = 5.6$, 7, and 5.4, respectively), whereas the fourth (C32) showed a low ratio of $R_{10:1} = 0.9$ (Fig. 2a). Interestingly, C32 is the active-site nucleophile of GSTO1 (ref. 22). Acetyl-CoA acetyltransferase-1 (ACAT1) was also labelled on four cysteines and three showed high ratios (C119, C196 and C413 showed ratios of $R_{10:1} = 8.8$, 8.2 and 4, respectively), whereas the fourth, the active site nucleophile C126 (ref. 23), yielded a low ratio of $R_{10:1} = 1.1$ (Fig. 2a).

The aforementioned findings indicated that heightened IA reactivity might be a good predictor of cysteine functionality in proteins. To examine this premise more systematically, we queried the Universal Protein Resource (UniProt) database to retrieve functional annotations for the 1,082 cysteine residues labelled by the IA probe. This analysis revealed that the most hyper-reactive cysteines were remarkably enriched in functional residues, with 35% of the cysteines with $R_{10:1} < 2$ being annotated as active-site nucleophiles or redox-active disulphides compared to 0.2% for all cysteine residues in the UniProt database (Fig. 2b, c, Supplementary Fig. 9 and Supplementary Tables 4 and 5). Hyper-reactive cysteines were also, as a group, more conserved across eukaryotic evolution (Supplementary Fig. 10). A broader survey of hyper-reactive cysteines identified several that have been ascribed functional properties in the literature despite lacking annotation in UniProt (Supplementary Fig. 11). For example, a single hyper-reactive cysteine C108 ($R_{10:1} = 1.0$) was identified in the uncharacterized protein D15Wsu75e. This protein and its orthologues are predicted

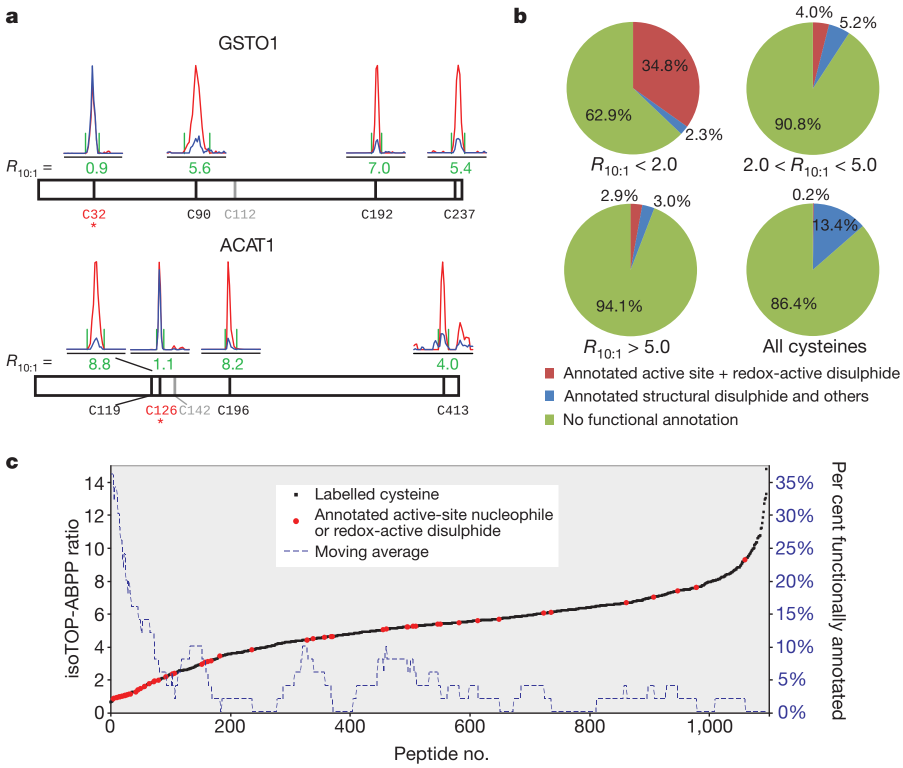


Figure 2 | Hyper-reactive cysteines are highly enriched in functional residues. a, Chromatographs from an isoTOP-ABPP experiment using 100:10 μM IA probe are shown for peptides from GSTO1 (top) and ACAT1 (bottom). The cysteine nucleophiles (asterisks) show low ratios ($R_{10:1} \approx 1$), whereas other cysteines show high ratios ($R_{10:1} \geq 4$). **b**, Pie charts illustrating the percentage of functionally annotated cysteines for three isoTOP-ABPP ratio ranges, including an average derived from all cysteines in the UniProt database. **c**, Correlation of isoTOP-ABPP ratios with functional annotations from the UniProt database where active-site nucleophiles or redox-active disulphides are shown in red, and all other cysteines in black. A moving average (window of 50) of functional residues is shown as a dashed blue line, demonstrating a profound enrichment within $R_{10:1} < 2.0$. Data are from experiments in three human cancer cell lines (MCF7, MDA-MB-231 and Jurkat).

to be cysteine proteases based on conservation of a prototypical Cys-His catalytic dyad[24]. Interestingly, C108 corresponds to the putative cysteine nucleophile of this catalytic motif and a recent crystal structure confirms the proximity of C108 to a conserved histidine (H38) (Supplementary Fig. 12). Thus, quantitative reactivity profiling supports structural predictions that D15Wsu75e is a functional cysteine protease.

Hyper-reactive cysteines also corresponded to sites for post-translational modification. For instance, C101 ($R_{10:1} = 1.92$) in the protein arginine methyltransferase PRMT1 has been identified as a site of modification by the endogenous oxidative product 4-hydroxy-2-nonenal (HNE)[25]. This cysteine, although nonessential for catalytic function, is an active site residue that makes direct contact with the *S*-adenosylmethionine cofactor[26] (Fig. 3a). Interestingly, we found that HNE inhibited both the IA-labelling (Fig. 3b) and catalytic activity (Fig. 3c) of wild-type PRMT1. A C101A mutant of PRMT1 showed substantially reduced IA-labelling (Fig. 3b) and HNE sensitivity (Fig. 3c). These data indicate that PRMT1 may be regulated by oxidative stress pathways through selective HNE modification of its hyper-reactive, active-site C101 residue. Additional hyper-reactive cysteines represented sites for glutathionylation[27] (CLIC1 (C24), CLIC3 (C25) and CLIC4 (C35); $R_{10:1} = 2.02$, 1.07 and 1.45, respectively) and nitrosylation[28] (RTN3; C42, $R_{10:1} = 0.78$). These data, taken together, indicate that heightened reactivity is not only a feature of catalytic cysteines, but also of 'non-catalytic', active-site cysteines, as well as those that undergo various forms of oxidative modification.

Function of the hyper-reactive cysteine in FAM96B

Intrigued by the diverse functional properties showed by hyper-reactive cysteines, we reasoned that critical activities might be inferred for such residues in hitherto uncharacterized proteins. A survey of the cysteines displaying low isoTOP-ABPP ratios uncovered the highly conserved C93 ($R_{10:1} = 1.15$) in the uncharacterized protein FAM96B (Supplementary Fig. 13). FAM96B has close orthologues in many organisms including the YHR122W protein from the budding yeast *Saccharomyces cerevisiae*, which shows 52% identity with human FAM96B, including conservation of C93 (the corresponding residue in YHR122W is C161). The gene encoding YHR122W is essential for yeast viability[29], and we found that expression of wild-type YHR122W, but not the C161A mutant of YHR122W could rescue a yeast strain in which the *YHR122W* gene was conditionally suppressed (Fig. 4a and

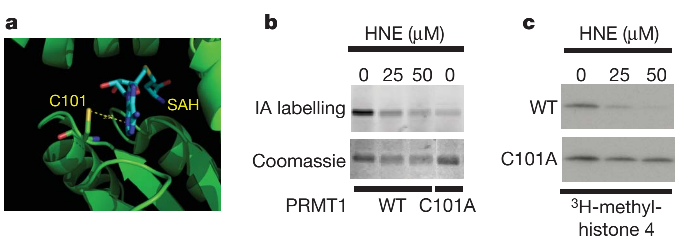


Figure 3 | Functional characterization of the hyper-reactive cysteines in PRMT1. a, Crystal structure of rat PRMT1[26] (green, PDB accession code 1ORI) showing the hyper-reactive cysteine C101 in contact with an *S*-adenosylhomocysteine (SAH) cofactor (cyan). **b**, Wild type (WT) and C101A mutant of human PRMT1 were labelled with the IA probe, followed by click chemistry to incorporate a fluorescent rhodamine tag. In-gel fluorescence demonstrates robust labelling of the wild-type but not C101A mutant PRMT1, and shows that IA-probe labelling of wild-type PRMT1 is inhibited by HNE (upper panel). Lower panel shows Coomassie blue staining for treated protein samples. **c**, Catalytic activity of purified wild-type, but not C101A mutant PRMT1 is inhibited by HNE as measured by monitoring transfer of ^{3}H-methyl from ^{3}H-*S*-adenosylmethionine (SAM) to a histone 4 substrate.

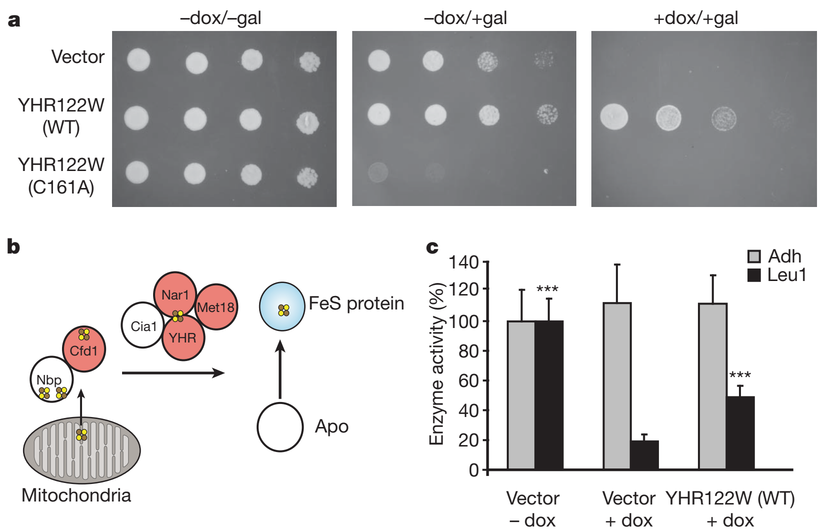


Figure 4 | Functional characterization of YHR122W/FAM96B. **a**, Expression of wild type and a C161A mutant of YHR122W in a yeast strain with a doxycycline (dox)-repressable *YHR122W* gene demonstrated a dominant-negative phenotype on induction of the C161A mutant expression (−dox/+gal, middle panel) and rescue of viability by expression of wild type, but not the C161A mutant of YHR122W (+dox/+gal, right panel). **b**, The cytosolic FeS cluster assembly pathway contains multiple proteins with hyper-reactive cysteines (in red). YHR122W/FAM96B (YHR) is a putative member of this network based on protein–protein interaction studies (see http://www.yeastgenome.org/). This panel was adapted from ref. 30. **c**, Doxycycline treatment of the *YHR122W*-repressable yeast strain significantly decreased the activity of the cytosolic FeS enzyme Leu1[41], and this activity is rescued by overexpression of wild-type YHR122W. These treatments had no effect on the activity of the non-FeS enzyme alcohol dehydrogenase (ADH). Error bars represent standard deviation, $n = 3$. ***$P < 0.001$, Student's t-test.

Supplementary Fig. 14). These data confirm the importance of C161 for the *in vivo* function of YHR122W and, by extension, other members of the FAM96B family.

We also observed that expression of the C161A mutant of YHR122W caused a severe growth defect in non-suppressive media indicative of a dominant-negative phenotype (Fig. 4a and Supplementary Fig. 14). This result indicates that the YHR122W protein may engage in protein complexes that are sequestered by the C161A mutant, thereby disrupting the activity of the wild-type protein. Consistent with this premise, queries of the *Saccharomyces* genome databank (SGD) revealed that YHR122W has been found in several large-scale protein interaction studies to bind to proteins involved in cytosolic iron-sulphur (FeS) cluster assembly, namely Nar1 and Cia1 (ref. 30; Fig. 4b). We found that the activity of the FeS-client protein isopropylmalate isomerase (Leu1)[31] was markedly reduced in YHR122W-deleted yeast, and this reduction was substantially rescued by expression of the wild-type YHR122W protein (Fig. 4c). These data support a role for the YHR122W/FAM96B protein in FeS-protein biogenesis. We also note that reactive cysteines seem to be a common feature of proteins in the FeS-protein assembly complex, including the human orthologues of Nar1, Met18 and Cfd1 (NARF, MMS19 and NUBP2, respectively) ($R_{10:1} = 0.91$, 2.2 and 2.9 respectively) (Supplementary Fig. 11), where they may assist in the transfer of assembled FeS clusters to client proteins[32].

Predicting functional cysteines in designed proteins

The marked correlation between cysteine hyper-reactivity and functionality observed in native proteomes led us to ask whether this relationship would extend to *de novo* designed proteins. We compared the IA labelling of twelve proteins that were computationally designed to act as cysteine hydrolases. These proteins originated from structurally distinct scaffolds and were all designed to contain cysteine-histidine dyads within an active site cavity (see Supplementary Methods for more details). Two of the designed proteins, ECH13 and ECH19, showed significant hydrolytic activity using a fluorogenic ester substrate, whereas the other ten designs were inactive (Fig. 5a and Supplementary Fig. 15a).

We first evaluated IA labelling of protein designs using a clickable, fluorescent reporter tag and SDS–polyacrylamide gel electrophoresis (SDS–PAGE) analysis, where similar amounts of each protein were tested in a homogeneous background proteome representing a mix of *Escherichia coli* and human (MCF7 cell line) proteins. The two active protein designs ECH13 and ECH19 showed strong IA-labelling signals compared to inactive designs (Fig. 5a), and, in both cases, mutation of the active-site cysteine to alanine abolished labelling (Fig. 5b) and hydrolytic activity (data not shown). We next combined the proteomes containing all twelve protein designs, diluted them into a background human cell proteome, and analysed the mixture by isoTOP-ABPP. Notably, both ECH13 and ECH19 showed isoTOP-ABPP ratios that were equivalent to the most hyper-reactive cysteines in human and *E. coli* proteomes ($R_{10:1} = 0.92$ and 1.27, respectively), whereas the remaining inactive protein designs all showed higher ratios ranging from 1.88–6.11 (Fig. 5c and Supplementary Fig. 15b, c). These data thus reveal a strong correlation between cysteine hyper-reactivity and hydrolytic activity across a diverse panel of protein designs and designate heightened cysteine nucleophilicity as a key feature of successful cysteine hydrolase designs.

Conclusions

Here, we have described a quantitative method to profile the intrinsic reactivity of cysteine residues in native proteomes. Measurement of the rate of alkylation by IA (or other carbon electrophiles) has been used by enzymologists to assess the nucleophilicity of cysteine residues in individual, purified proteins[6]. With isoTOP-ABPP, these studies can now be extended to quantitative, proteome-wide surveys of cysteine reactivity in complex biological systems. A key advantage of isoTOP-ABPP over more traditional proteomic methods that target cysteine-containing peptides[14,18] is the use of an alkynylated IA probe in place of more bulky biotinylated reagents, which have shown an impaired ability to label cysteines in native proteins[14]. Alkynylated IA probes, owing to their cell permeability, also afford the opportunity to perform cysteine reactivity profiling in living systems. In pilot experiments, we have found that a large fraction of hyper-reactive cysteines are labelled by the IA probe in living cells (Supplementary Fig. 16). Furthermore, isoTOP-ABPP selectively targets probe-accessible cysteines in native proteins. In this way, structural cysteines engaged in disulphide bonds or buried within the body of a protein are avoided to provide preferential access to a specific fraction of cysteines that are profoundly enriched in functionality (the

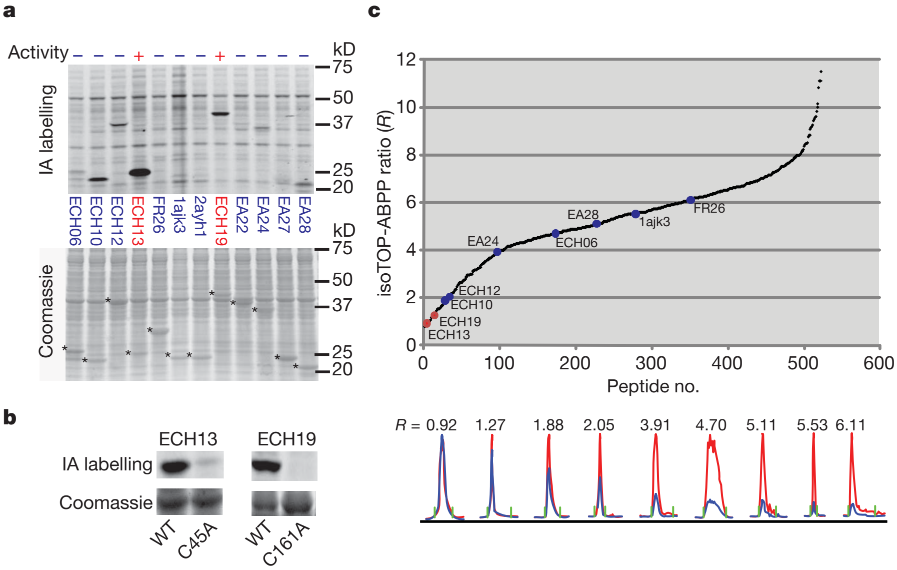


Figure 5 | Quantitative reactivity profiling predicts functional cysteines in designed proteins. **a**, In-gel fluorescence demonstrates robust IA labelling of two active cysteine hydrolases, ECH13 and ECH19, relative to inactive designs (top panel). Hydrolysis activities of ECH13 and ECH19 measured as the ratio of velocities in the presence versus the absence of purified enzymes were 71.64 ± 6.94 and 104.15 ± 10.78, respectively (see Supplementary Fig. 15a for substrate hydrolysis assay). Other designs showed no measurable hydrolysis activity over background (0.76 ± 0.058 nmol s^{-1}). Asterisks designate Coomassie blue signals for protein designs (lower panel). **b**, IA labelling is observed for ECH13 and ECH19, but not their active-site cysteine mutants C45A and C161A, respectively. **c**, Catalytic cysteines in ECH13 and ECH19 show low isoTOP-ABPP ratios (red) compared with other designs (blue). Chromatographs are shown for peptides from the nine designs identified in this experiment (bottom panel), in the same order as shown in the top panel.

IA probe labelled 1,082 out of a total of 8,910 cysteines present on the 890 human proteins detected in this study). Projecting forward, it is possible that, by varying the nature of the electrophile, isoTOP-ABPP probes can be created that profile the reactivity of different subsets of cysteines, as well as other amino acids in proteomes, such as serine, threonine, tyrosine and glutamate/aspartate, which have also been shown to react with small-molecule probes[16,18,33–35].

We discovered that hyper-reactivity can predict cysteine function in both native and designed proteins. The fact that hyper-reactivity was strongly correlated with catalytic activity in *de novo* designed cysteine hydrolases is interesting from the principles of both enzyme engineering and assay development, as it indicates that heightened cysteine nucleophilicity is a key feature of active catalysts and, accordingly, electrophile reactivity could serve as an effective primary screen for novel cysteine-dependent enzymes. We show that these screens can be performed directly in complex proteomes using either gel or MS (isoTOP-ABPP) detection platforms, thus offering a versatile and relatively high-throughput way to evaluate many protein designs in parallel. The isoTOP-ABPP platform has the additional advantage of reading out the relative cysteine reactivity of designs independent of their expression levels against a ‘background’ of native, hyper-reactive cysteines for comparison. isoTOP-ABPP might also offer a complementary way to perform cysteine reactivity/accessibility experiments that monitor protein stability and ligand interactions[36,37].

The relationship between cysteine reactivity and functionality extends beyond nucleophilic catalysis to include other enzymatic activities (oxidative/reductive), as well as sites of electrophilic and oxidative modification. Quantitative reactivity profiling thus distinguishes itself as a complementary and perhaps more inclusive strategy to survey cysteine function compared to previous computational[9] and experimental[11–14,17] methods that focus on specific cysteine-based activities or modification events. Considering further that hyper-reactive cysteines corresponded to sites for glutathionylation[27], nitrosylation[28] and HNE-modification[25], we speculate that cysteine nucleophilicity is a property that may have been selected for during evolution to offer points of protein control by oxidative stress pathways. Determining how the reactivity of cysteine residues is honed will require further investigation, but we anticipate that quantitative proteomic data, when integrated with the output of ongoing structural genomics programs, may eventually uncover unifying mechanistic principles that explain cysteine reactivity in proteins. In this regard, it is interesting to note that, although hyper-reactive cysteines did not conform to any obvious consensus sequence motifs, many of these residues were found at the N termini of α-helices (Supplementary Fig. 17). This finding is consistent with literature reports ascribing a role for α-helix dipoles in the stabilization of cysteine thiolate anions[38].

Finally, it is important to stress that some functional cysteines may be inherently reactive, but inaccessible to our IA probe for steric reasons. Other cysteine-reactive electrophilic probes[16,17] may prove more suitable for such cysteine residues. Also, hyper-reactivity is not necessarily a defining feature for all functional cysteines. Some enzymes with catalytic cysteines may, for instance, show reduced reactivity until they bind their physiological substrates or may rely more on substrate recognition than inherent catalytic power for function. This may be the case with the E1-activating and E2-conjugating enzymes, which recognize a specific class of ubiquitinated substrates and possess active-site cysteines that showed only moderate levels of electrophile reactivity (Supplementary Fig. 18). Other cysteines may have activities that are not dependent on their nucleophilicity. Our data do indicate, however, that those cysteines that are hyper-reactive in proteomes probably perform important catalytic and/or regulatory functions for their parent proteins. The large number of newly discovered residues that fall into this category foretell a broad role for hyper-reactive cysteines in mammalian biology.

METHODS SUMMARY

Probes and tags. The IA probe and the light and heavy variants of the azide-TEV-biotin tags were synthesized as previously described[20,39].

Sample preparation, mass spectrometry and data analysis. For concentration-dependent experiments, proteome samples in PBS were probe labelled with the desired probe concentration for 1 h. Click chemistry was performed with either the light or heavy variants of the azide-TEV-biotin tags and the samples were mixed and subjected to streptavidin enrichment and subsequent trypsin and TEV digestion. The resulting TEV digests were analysed by Multidimensional Protein Identification Technology (MudPIT) on an LTQ-Orbitrap instrument. The resulting tandem MS data were searched using the SEQUEST algorithm[40] using

a concatenated target/decoy variant of the human, mouse and *E. coli* protein sequence databases. Quantification of light:heavy ratios (isoTOP-ABPP ratios, *R*) was performed using in-house software. Detailed information on sample preparation, mass spectrometry methods and data analysis is presented in Methods.

Complementation of *S. cerevisiae* YHR122W deletion mutant. Complementary DNA encoding wild-type YHR122W was subcloned into the pESC_Leu vector (Stratagene). The YHR122W(C161A) mutant was generated using the Quickchange procedure (Stratagene). These constructs were introduced into a yeast Tet promoter Hughes (yTHC) strain harbouring a conditional (doxycycline-dependent) disruption in the *YHR122W* gene (Open Biosystems). Growth of these transformed cell lines on ±gal/±dox media was monitored for 3 days. These cell lines were also used to monitor Leu1 and alcohol dehydrogenase (ADH) activity. Detailed information on the protocols used to subclone, transform and monitor the growth of the yeast strains and measure enzyme activity is available in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 May; accepted 3 September 2010.
Published online 17 November 2010.

1. Eisenberg, D., Marcotte, E. M., Xenarios, I. & Yeates, T. O. Protein function in the post-genomic era. *Nature* **405,** 823–826 (2000).
2. Zhao, Y. & Jensen, O. N. Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques. *Proteomics* **9,** 4632–4641 (2009).
3. Bulaj, G., Kortemme, T. & Goldenberg, D. P. Ionization reactivity relationships for cysteine thiols in polypeptides. *Biochemistry* **37,** 8965–8972 (1998).
4. Giles, N. M., Giles, G. I. & Jacob, C. Multiple roles of cysteine in biocatalysis. *Biochem. Biophys. Res. Commun.* **300,** 1–4 (2003).
5. Reddie, K. G. & Carroll, K. S. Expanding the functional diversity of proteins through cysteine oxidation. *Curr. Opin. Chem. Biol.* **12,** 746–754 (2008).
6. Voss, A. A., Lango, J., Ernst-Russell, M., Morin, D. & Pessah, I. N. Identification of hyperreactive cysteines within ryanodine receptor type 1 by mass spectrometry. *J. Biol. Chem.* **279,** 34514–34520 (2004).
7. Lewis, C. T., Seyer, J. M. & Carlson, G. M. Cysteine 288: an essential hyperreactive thiol of cytosolic phosphoenolpyruvate carboxykinase (GTP). *J. Biol. Chem.* **264,** 27–33 (1989).
8. Knowles, J. R. Intrinsic pK_a-values of functional-groups in enzymes: improper deductions from pH-dependence of steady-state parameters. *CRC Crit. Rev. Biochem.* **4,** 165–173 (1976).
9. Fomenko, D. E., Xing, W., Adair, B. M., Thomas, D. J. & Gladyshev, V. N. High-throughput identification of catalytic redox-active cysteine residues. *Science* **315,** 387–389 (2007).
10. Sethuraman, M. *et al.* Isotope-coded affinity tag (ICAT) approach to redox proteomics: identification and quantitation of oxidant-sensitive cysteine thiols in complex protein mixtures. *J. Proteome Res.* **3,** 1228–1233 (2004).
11. Baty, J. W., Hampton, M. B. & Winterbourn, C. C. Proteomic detection of hydrogen peroxide-sensitive thiol proteins in Jurkat cells. *Biochem. J.* **389,** 785–795 (2005).
12. Salsbury, F. R. Jr, Knutson, S. T., Poole, L. B. & Fetrow, J. S. Functional site profiling and electrostatic analysis of cysteines modifiable to cysteine sulfenic acid. *Protein Sci.* **17,** 299–312 (2008).
13. Leonard, S. E., Reddie, K. G. & Carroll, K. S. Mining the thiol proteome for sulfenic acid modifications reveals new targets for oxidation in cells. *ACS Chem. Biol.* **4,** 783–799 (2009).
14. Kim, J.-R., Yoon, H. W., Kwon, K.-S., Lee, S.-R. & Rhee, S. G. Identification of proteins containing cysteine residues that are sensitive to oxidation by hydrogen peroxide at neutral pH. *Anal. Biochem.* **283,** 214–221 (2000).
15. Speers, A. E. & Cravatt, B. F. A tandem orthogonal proteolysis strategy for high-content chemical proteomics. *J. Am. Chem. Soc.* **127,** 10018–10019 (2005).
16. Weerapana, E., Simon, G. M. & Cravatt, B. F. Disparate proteome reactivity profiles of carbon electrophiles. *Nature Chem. Biol.* **4,** 405–407 (2008).
17. Dennehy, M. K., Richards, K. A., Wernke, G. R., Shyr, Y. & Liebler, D. C. Cytosolic and nuclear protein targets of thiol-reactive electrophiles. *Chem. Res. Toxicol.* **19,** 20–29 (2006).
18. Shin, N.-Y., Liu, Q., Stamer, S. L. & Liebler, D. C. Protein targets of reactive electrophiles in human liver microsomes. *Chem. Res. Toxicol.* **20,** 859–867 (2007).
19. Speers, A. E., Adam, G. C. & Cravatt, B. F. Activity-based protein profiling *in vivo* using a copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Am. Chem. Soc.* **125,** 4686–4687 (2003).
20. Weerapana, E., Speers, A. E. & Cravatt, B. F. Tandem orthogonal proteolysis-activity-based protein profiling (TOP-ABPP)—a general method for mapping sites of probe modification in proteomes. *Nature Protocols* **2,** 1414–1425 (2007).
21. Shiio, Y. & Aebersold, R. Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry. *Nature Protocols* **1,** 139–145 (2006).
22. Board, P. G. *et al.* Identification, characterization, and crystal structure of the omega class glutathione transferases. *J. Biol. Chem.* **275,** 24798–24806 (2000).
23. Thompson, S. *et al.* Mechanistic studies on β-ketoacyl thiolase from *Zoogloea ramigera*: identification of the active-site nucleophile as Cys89, its mutation to Ser89, and kinetic and thermodynamic characterization of wild-type and mutant enzymes. *Biochemistry* **28,** 5735–5742 (1989).
24. Iyer, L. M., Koonin, E. V. & Aravind, L. Novel predicted peptidases with a potential role in the ubiquitin signaling pathway. *Cell Cycle* **3,** 1440–1450 (2004).
25. Codreanu, S. G., Zhang, B., Sobecki, S. M., Billheimer, D. D. & Liebler, D. C. Global analysis of protein damage by the lipid electrophile 4-hydroxy-2-nonenal. *Mol. Cell. Proteomics* **8,** 670–680 (2009).
26. Zhang, X. & Cheng, X. Structure of the predominant protein arginine methyltransferase PRMT1 and analysis of Its binding to substrate peptides. *Structure* **11,** 509–520 (2003).
27. Harrop, S. J. *et al.* Crystal structure of a soluble form of the intracellular chloride ion channel CLIC1 (NCC27) at 1.4-Å resolution. *J. Biol. Chem.* **276,** 44993–45000 (2001).
28. Hao, G., Derakhshan, B., Shi, L., Campagne, F. & Gross, S. S. SNOSID, a proteomic method for identification of cysteine S-nitrosylation sites in complex protein mixtures. *Proc. Natl Acad. Sci. USA* **103,** 1012–1017 (2006).
29. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418,** 387–391 (2002).
30. Lill, R. Function and biogenesis of iron sulphur proteins. *Nature* **460,** 831–838 (2009).
31. Pierik, A. J., Netz, D. J. & Lill, R. Analysis of iron–sulfur protein maturation in eukaryotes. *Nature Protocols* **4,** 753–766 (2009).
32. Netz, D. J. A., Pierik, A. J., Stumpfig, M., Muhlenhoff, U. & Lill, R. The Cfd1–Nbp35 complex acts as a scaffold for iron-sulfur protein assembly in the yeast cytosol. *Nature Chem. Biol.* **3,** 278–286 (2007).
33. Okerberg, E. S. *et al.* High-resolution functional proteomics by active-site peptide profiling. *Proc. Natl Acad. Sci. USA* **102,** 4996–5001 (2005).
34. Nazif, T. & Bogyo, M. Global analysis of proteasomal substrate specificity using positional-scanning libraries of covalent inhibitors. *Proc. Natl Acad. Sci. USA* **98,** 2967–2972 (2001).
35. Chen, G. *et al.* Reactivity of functional groups on the protein surface: development of epoxide probes for protein labeling. *J. Am. Chem. Soc.* **125,** 8130–8133 (2003).
36. Silverman, J. A. & Harbury, P. B. Rapid mapping of protein structure, interactions, and ligand binding by misincorporation proton-alkyl exchange. *J. Biol. Chem.* **277,** 30968–30975 (2002).
37. Isom, D. G., Vardy, E., Oas, T. G. & Hellinga, H. W. Picomole-scale characterization of protein stability and function by quantitative cysteine reactivity. *Proc. Natl Acad. Sci. USA* **107,** 4908–4913 (2010).
38. Kortemme, T. & Creighton, T. E. Ionisation of cysteine residues at the termini of model α-helical peptides. Relevance to unusual thiol pK_a values in proteins of the thioredoxin family. *J. Mol. Biol.* **253,** 799–812 (1995).
39. Macpherson, L. J. *et al.* Noxious compounds activate TRPA1 ion channels through covalent modification of cysteines. *Nature* **445,** 541–545 (2007).
40. Eng, J. K., Mccormack, A. L. & Yates, J. R. An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5,** 976–989 (1994).
41. Kispal, G., Csere, P., Prohl, C. & Lill, R. The mitochondrial proteins Atm1p and Nfs1p are essential for biogenesis of cytosolic Fe/S proteins. *EMBO J.* **18,** 3981–3989 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank T. Bartfai, I. Wilson and members of the B.F.C. laboratory for comments and critical reading of the manuscript, T. Ji for experimental assistance and J. Gallaher for expression of designed proteins. This work was supported by the National Institutes of Health (CA087660, MH084512), a Pfizer Postdoctoral Fellowship (E.W.), a Koshland Graduate Fellowship in Enzyme Biochemistry (G.M.S.), a National Science Foundation predoctoral fellowship (D.A.B.) and the Skaggs Institute for Chemical Biology.

Author Contributions B.F.C., E.W. and C.W. conceived the project and E.W. and C.W. performed MS experiments and yeast growth/Leu1 assays. C.W. and G.M.S. performed computational data analyses. S.K., F.R. and D.B. performed computational design of cysteine hydrolases and measured activity using a fluorogenic assay. D.A.B. purified PRMT1 and M.B.D.D. and K.M. performed PRMT1 activity assays. B.F.C., E.W., C.W. and G.M.S. analysed data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.F.C. (cravatt@scripps.edu).

METHODS

All compounds and reagents were purchased from Novabiochem, Sigma or Fisher, except where noted.

Preparation of mouse proteomes. Mouse tissues (heart and liver) were harvested and immediately flash frozen in liquid nitrogen. The tissues were then Dounce homogenized in 1× PBS, pH 7.4. Centrifugation at 100,000*g* (45 min) provided soluble fractions (supernatant) and membrane fractions (pellet). Protein concentrations for each proteome were obtained using the Bio-Rad DC protein assay and stored at −80 °C till use.

Preparation of human cancer cell line proteomes. MDA-MB-231 cells were grown in L15 media supplemented with 10% fetal bovine serum at 37 °C in a CO_2-free incubator. Jurkat cells and MCF7 cells were grown in RPMI-1640 supplemented with 10% fetal bovine serum at 37 °C with 5% CO_2. For *in vitro* labelling experiments, cells were grown to 100% confluency, washed three times with PBS and scraped in cold PBS. Cell pellets were isolated by centrifugation at 1,400*g* for 3 min, and the cell pellets stored at −80 °C until further use. For *in situ* labelling of MDA-MB-231 and MCF7 cells, the cells were grown to 90% confluency, the media was removed and replaced with fresh media containing 10 μM IA probe. The cells were incubated at 37 °C for 1 h and harvested as detailed above. The harvested cell pellets were lysed by sonication and fractionated by centrifugation (100,000*g*, 45 min) to yield soluble and membrane proteomes. The proteomes were diluted to 2 mg ml^{-1} and stored at −80 °C until use.

Protein labelling and click chemistry. Proteome samples were diluted to a 2 mg protein/ml solution in PBS. Each sample (2 × 0.5 ml aliquots) was treated with 10, 20, 50, or 100 μM of IA probe using 5 μl of a 1, 2, 5, or 10 mM stock in DMSO. The labelling reactions were incubated at room temperature (25 °C) for 1 h. Click chemistry was performed by the addition of 150 μM of either the light TEV tag or heavy TEV tag (15 μl of a 5 mM stock), 1 mM tris(2-carboxyethyl)phosphine (TCEP; fresh 50× stock in water), 100 μM ligand (17× stock in DMSO:t-butanol 1:4) and 1 mM $CuSO_4$ (50× stock in water). Samples were allowed to react at room temperature for 1 h. After the click chemistry step, the light- and heavy-labelled samples were mixed together and centrifuged (5,900*g*, 4 min, 4 °C) to pellet the precipitated proteins. The pellets were washed twice in cold MeOH, after which the pellet was solubilized in PBS containing 1.2% SDS via sonication and heating (5 min, 80 °C).

For time course experiments, proteome samples were labelled with 100 μM of IA probe (using 5 μl of a 10 mM stock in DMSO). After 6 min of probe labelling, an aliquot of the reaction was quenched by passaging the sample through a NAP-5 column (GE Healthcare) to remove excess, unreacted probe. After 60 min of probe labelling, the other sample was quenched as before and click chemistry was performed as described earlier.

Streptavidin enrichment of probe-labelled proteins. The SDS-solubilized, probe-labelled proteome samples were diluted with 5 ml of PBS for a final SDS concentration of 0.2%. The solutions were then incubated with 100 μl of streptavidin-agarose beads (Pierce) for 3 h at room temperature. The beads were washed with 10 ml 0.2% SDS/PBS, 3 × 10 ml PBS and 3 × 10 ml H_2O and the beads were pelleted by centrifugation (1,300*g*, 2 min) between washes.

On-bead trypsin and TEV digestion. The washed beads described earlier were suspended in 500 μl of 6 M urea/PBS and 10 mM TCEP (from 20× stock in H_2O) and placed in a 65 °C heat block for 15 min. Twenty millimolar iodoacetamide (from 50× stock in H_2O) was then added and allowed to react at 37 °C for 30 min. Following reduction and alkylation, the beads were pelleted by centrifugation (1,300*g*, 2 min) and resuspended in 200 μl of 2 M urea/PBS, 1 mM $CaCl_2$ (100× stock in H_2O), and trypsin (2 μg). The digestion was allowed to proceed overnight at 37 °C. The digest was separated from the beads using a Micro Bio-Spin column and the beads were then washed with 3 × 500 μl PBS, 3 × 500 μl H_2O, and 1 × 150 μl of TEV digest buffer. The washed beads were then resuspended in 150 μl of TEV digest buffer with AcTEV Protease (Invitrogen, 5 μl) for 12 h at 29 °C. The eluted peptides were separated from the beads using a Micro Bio-Spin column and the beads washed with H_2O (2 × 75 μl). Formic acid (15 μl) was added to the sample, which was stored at −20 °C until MS analysis.

Liquid-chromatography-mass-spectrometry (LC-MS) analysis. LC-MS/MS analysis was performed on an LTQ-Orbitrap mass spectrometer (ThermoFisher) coupled to an Agilent 1100 series high-performance liquid chromatography system. TEV digests were pressure loaded onto a 250 μm fused silica desalting column packed with 4 cm of Aqua C18 reverse phase resin (Phenomenex). The peptides were then eluted onto a biphasic column (100 μm fused silica with a 5 μm tip, packed with 10 cm C18 and 3 cm Partisphere strong cation exchange resin (SCX, Whatman) using a gradient 5–100% buffer B in buffer A (buffer A: 95% water, 5% acetonitrile, 0.1% formic acid; buffer B: 20% water, 80% acetonitrile, 0.1% formic acid). The peptides were then eluted from the SCX onto the C18 resin and into the mass spectrometer using four salt steps as previously described[15,20]. The flow rate through the column was set to ~0.25 μl min^{-1} and the spray voltage was set to 2.75 kV. One full MS scan (FTMS) (400–1,800 MW) was followed by 18 data dependent scans (ITMS) of the *n*th most intense ions with dynamic exclusion disabled.

Peptide identification. The tandem MS data were searched using the SEQUEST algorithm[40] using a concatenated target/decoy variant of the human and mouse International Protein Index databases. A static modification of +57.02146 on cysteine was specified to account for iodoacetamide alkylation and differential modifications of +464.28596 (light probe modification) and +470.29977 (heavy probe modification) were specified on cysteine to account for probe modifications with the either light or heavy variants of the IA-probe-TEV adduct. SEQUEST output files were filtered using DTASelect 2.0[42]. Reported peptides were required to be fully tryptic and contain the desired probe modification and discriminant analyses were performed to achieve a peptide false-positive rate below 5%. The actual false-positive rate was assessed at this stage according to established guidelines[43] and found to be ~3.5%. Additional assessments of the false-positive rate were performed following the application of additional filters (described later) resulting in a final false-positive rate below 0.05%.

Ratio quantification. Quantification of light/heavy ratios (isoTOP-ABPP ratios, *R*) was performed using in-house software written in the R programming language that utilizes routines from the open-source XCMS package[44] for MS data analysis to read in raw chromatographic data in the mzXML format[45]. Each experiment consisted of two LC/LC-MS/MS runs: light:heavy 10 μM:10 μM, and light:heavy 100 μM:10 μM IA-probe concentration. Both runs were searched using SEQUEST and filtered with DTASelect as described earlier. Because the mass spectrometer was configured for data-dependant fragmentation, peptides are not always identified in every run. As such, peptides were identified in either 1) only the 10 μM:10 μM run, 2) only the 100 μM:10 μM run, or 3) both runs. In the case of peptides that were sequenced in both runs, identification of the corresponding peaks was made by choosing peaks that co-elute with the peptide identification. In the case of probe-modified peptides that were sequenced in one, but not the other run, an algorithm was developed to identify the corresponding peak in the run without the SEQUEST identification. To accomplish this, the retention time of the 'reference' peptide is used to position a retention time window (± 10 min) across the run lacking a peptide identification. Extracted ion chromatograms (± 10 p.p.m.) of the target peptide *m/z* with both 'light' and 'heavy' modifications are generated within that window. The program then searches for candidate co-eluting pairs of light:heavy MS1 peaks, and for each candidate pair calculates the ratio of integrated peak area between the light and heavy peaks. Several filters are used to ensure that the correct peak pair is identified. First, the extent of co-elution for each peak pair is quantified using a Pearson correlation, an established method to gauge elution profile similarity[46]. Second, the predicted pattern of the isotopic envelope of the target peptide is generated and compared to the observed high-resolution MS1 spectrum. This comparison generates an 'envelope correlation score' (Env) that also enables confirmation of the monoisotopic mass and charge state of each candidate peak. Peak pairs that have poor co-elution scores, or that have the incorrect monoisotopic mass or charge, or whose isotopic envelopes are not well correlated with the predicted envelope are eliminated from consideration. After application of these filters, in the rare case that multiple candidates still exist, then no peak is chosen and a ratio is not recorded. Usually, however, application of these filters results in a single candidate peak pair and the ratio for this peak pair is recorded for the peptide in the corresponding run. In this way, each experiment yields two ratios, one for the 10 μM:10 μM run and one for the 100 μM:10 μM run. Following application of these filters, the false-positive rate was reassessed, and found to be less than 0.05% in all cases.

After ratios for unique peptide entries are calculated for each experiment, overlapping peptides with the same labelled cysteine (for example, same local sequence around the labelled cysteines but different charge states, MudPIT segment numbers, or tryptic termini) are grouped together, and the median ratio from each group is reported as the final ratio (*R*). All of these values can be found in Supplementary Tables 1, 2 and 3 and representative chromatographs can be seen in Supplementary Table 7. Raw result files of peptide identification using SEQUEST can be found in Supplementary Table 9.

Functional annotation of labelled cysteines. For automated functional analyses, custom perl-scripts were developed to query the UniProtKB/Swiss-Prot Protein Knowledgebase release 57.4 (current as of 16 June 2009). Sequence annotation in the (Features) section of the relevant UniProt entry was mined and any annotation corresponding to the labelled residue was collected. This functional annotation in its entirety can be found in Supplementary Tables 4 and 5.

Recombinant PRMT1 protein expression and purification. Full-length cDNA encoding human PRMT1 in pOTB7 was purchased from Open BioSystems and subcloned into pET-45b(+) (Novagen). BL21(DE3) *E. coli* containing this vector was grown in LB media containing 75 mg l^{-1} carbenicillin with shaking at 37 °C

to an $OD_{600\,nm}$ of 0.5. The cells were then induced with 1 mM isopropyl-β-D-thiogalactoside (IPTG) and harvested 4 h later by centrifugation. Cells were lysed by stirring for 20 min at 4 °C in 50 mM Tris-HCl (pH 8.0) with 150 mM NaCl and supplemented with 1 mg ml^{-1} lysozyme and 1 mg ml^{-1} DNase I. The lysate was then sonicated and centrifuged at 10,000*g* for 10 min. Talon cobalt affinity resin (Clontech; 400 μl of slurry per gram of cell paste) was added to the supernatant, and the mixture was rotated at 25 °C for 30 min. Beads were collected by centrifugation at 700*g* for 3 min, washed twice with Tris buffer, and applied to a 1-cm column. The column was washed twice with Tris buffer (10 ml per 400 μl of resin slurry) and Tris buffer with 500 mM NaCl once. The bound protein was eluted by the addition of 100 mM imidazole (2 ml per 400 μl of resin). Imidazole was removed by passage over a Sephadex G-25M column (GE Healthcare), and the eluate was concentrated using an Amicon centrifugal filter device (Millipore). Protein concentration was determined using the Bio-Rad DC protein assay kit. These conditions yielded PRMT1 at approximately 0.5 mg l^{-1} of culture. A C101A mutation was introduced into the pET-45b(+) construct described earlier using the Quikchange Site-Directed Mutagenesis Kit (Stratagene), and the resulting mutant protein was expressed identically and isolated with a similar yield.

In-gel fluorescence characterization of PRMT1. Thirteen micrograms of recombinant PRMT1 (wild type or C101A mutant) in 50 μl PBS buffer was pre-incubated with 0, 25 or 50 μM HNE (Calbiochem, 50 mM stock in ethanol) for 1 h at room temperature and was then labelled with 100 nM of the IA probe (5 μM stock in DMSO) and the reactions incubated for 1 h at room temperature. Click chemistry was performed with 20 μM rhodamine-azide, 1 mM TCEP, 100 μM TBTA ligand and 1 mM $CuSO_4$. The reaction was allowed to proceed at room temperature for 1 h before quenching with 50 μl of 2× SDS–PAGE loading buffer (reducing). Quenched reactions were separated by SDS–PAGE (30 μl of sample/lane) and visualized in-gel using a Hitachi FMBio IIe flatbed laser-induced fluorescence scanner (MiraiBio).

PRMT1 *in vitro* methylation assays. Five-hundred nanograms of recombinant human PRMT1 (wild type or C101A mutant) was pre-incubated with HNE (Calbiochem) for 30 min and methylation activity was monitored after addition of 1 mg of recombinant histone 4 (M2504S; NEB) and SAM (2 μCi) in methylation buffer (20 mM Tris, pH 8.0, 200 mM NaCl, 0.4 mM EDTA). Reactions were incubated for 90 min at 30 °C and stopped with SDS sample buffer. SDS–PAGE gels were fixed with 10% acetic acid/10% methanol v/v, washed, and incubated with Amplify reagent (Amersham) before exposing at −80 °C.

Complementation of *S. cerevisiae* YHR122W deletion mutant. A cDNA encoding YHR122W was purchased as a full-length expressed sequence tag (Open Biosystems). The construct for subcloning into the yeast epitope tagging vector pESC-Leu (Stratagene) was generated by polymerase chain reaction (PCR) from the corresponding cDNA using the following primers: sense primer, 5′-GAAGCGGCCGCAATGTCTGAGTTTTTGAATGA-3′; antisense primer, 5′- CCGACTAGTGCCTTACAAGTCACTAACATCTTAG-3′.

The PCR product was digested with NotI-SpeI and subcloned into a NotI-SpeI-digested pESC-Leu vector and sequenced. The YHR122W(C161A) mutant was generated using the Quickchange procedure (Stratagene). The mutant cDNA was sequenced and found to contain only the desired mutation.

Constructs containing wild-type and C161A mutant YHR122W were introduced into the yTHC strain YSC1180-7428770 (Open Biosystems) using the reagents provided in the Yeastmaker Yeast Transformation System 2 (Clontech). The yeast was grown in synthetic dextrose minimal medium (−Leu) and spot assays were performed in either synthetic dextrose minimal medium (−Leu) or synthetic galactose minimal medium (−Leu) + agar plates ± 50 μg ml^{-1} doxycycline. The plates were cultured at 30 °C for 3 days.

Isopropylmalate isomerase (Leu1) assay. Yeast strains harbouring either an empty vector or wild-type YHR122W (see earlier section) were cultured in synthetic dextrose minimal medium (−Leu) to an $OD_{600\,nm}$ of 1.0 and transferred into synthetic galactose minimal medium (−Leu) ± 50 μg ml^{-1} doxycycline for 12 h. Yeast were lysed and Leu1 semi-purified by ammonium sulphate precipitation (40–70%). The activity assays were performed using DL-threo-3-isopropylmalic acid as the substrate and product formation was measured by monitoring absorbance at 235 nm for 10 min[31].

ADH assay. Yeast cell lysates in 0.1 M sodium pyrophosphate buffer (pH 9.2, 1.5 ml) were treated with 2 M ethanol (0.5 ml) and 0.025 M NAD (1.0 ml) and ADH activity was measured by absorbance increase at 340 nm for 3 min[47].

***De novo* designs of cysteine hydrolases and hydrolysis activity assays.** We used the Rosetta computational enzyme design methodology[48] to search a set of protein scaffolds for constellations of backbones capable of supporting an idealized transition state for ester hydrolysis derived from the geometries and mechanisms of natural cysteine hydrolases[49]. The idealized active-site models feature a nucleophilic cysteine, a general base/acid histidine and at least one side-chain or backbone hydrogen bond donor as the oxyanion hole. The sequence of residues surrounding the putative active sites was optimized using the Rosetta design algorithm to maximize transition state stabilization[50]. A set of 12 designed proteins in 10 distinct scaffolds was chosen for experimental characterization. For each designed protein, synthetic genes were obtained and protein expression and purification was performed in *E. coli* as previously described[50]. Activity was measured with the substrate by following the initial (<5% substrate conversion) increase in fluorescence due to the appearance of the product coumarin. A protein concentration of 20 μM and substrate concentration of 100 μM were used in 25 mM HEPES buffer, 150 mM NaCl, 1 mM TCEP, pH 7.5. The background rate was measured under identical conditions but without the protein. Kunkel mutagenesis was used for creating point mutations in the active-site residues. A detailed description of the design and characterization of the cysteine hydrolases will be presented elsewhere. Amino acid sequences of the 12 designs can be found in Supplementary Information.

In-gel fluorescence and isoTOP-ABPP characterization of designed proteins. For in-gel fluorescence studies, *E.coli* lysates overexpressing the designed proteins were diluted to 2 mg protein/ml in PBS. Each sample (25 μl) was mixed with 25 μl of MCF7 human cell soluble proteome (2 mg ml^{-1}) and was labelled with 100 nM of the IA probe (5 μM stock in DMSO) and the reactions incubated for 1 h at room temperature. Click chemistry, SDS–PAGE separation and in-gel fluorescence visualization were performed as described in previous sections.

For isoTOP-ABPP studies, 10 ul of each of the *E.coli* lysates (2 mg protein/ml) overexpressing the designed constructs were mixed together and the total volume was brought to 1 ml by the addition of 2 mg ml^{-1} of MCF7 soluble proteome. Time-dependent and concentration-dependent labelling with the IA probe, click chemistry, on-bead trypsin and TEV digestions, LC-MS runs and MS data analysis were performed as described in previous sections.

42. Tabb, D. L., McDonald, W. H. & Yates, J. R. III. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **1,** 21–26 (2002).
43. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods* **4,** 207–214 (2007).
44. Collins, S. R. *et al.* Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics* **6,** 439–450 (2007).
45. Pedrioli, P. G. A. *et al.* A common open representation of mass spectrometry data and its application to proteomics research. *Nature Biotechnol.* **22,** 1459–1466 (2004).
46. Park, S. K., Venable, J. D., Xu, T. & Yates, J. R. A quantitative analysis software tool for mass spectrometry-based proteomics. *Nature Methods* **5,** 319–322 (2008).
47. Vallee, B. L. & Hoch, F. L. Zinc, a component of yeast alcohol dehydrogenase. *Proc. Natl Acad. Sci. USA* **41,** 327–338 (1955).
48. Zanghellini, A. *et al.* New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* **15,** 2785–2794 (2006).
49. Ma, S., Devi-Kesavan, L. S. & Gao, J. Molecular dynamics simulations of the catalytic pathway of a cysteine protease: a combined QM/MM study of human cathepsin K. *J. Am. Chem. Soc.* **129,** 13633–13645 (2007).
50. Jiang, L. *et al.* *De novo* computational design of retro-aldol enzymes. *Science* **319,** 1387–1391 (2008).

doi:10.1038/nature09601

A lower limit of $\Delta z > 0.06$ for the duration of the reionization epoch

Judd D. Bowman[1]* & Alan E. E. Rogers[2]*

Observations of the 21-centimetre line of atomic hydrogen in the early Universe directly probe the history of the reionization of the gas between galaxies[1]. The observations are challenging, though, because of the low expected signal strength (~10 mK), and contamination by strong (>100 K) foreground synchrotron emission in the Milky Way and extragalactic continuum sources[2]. If reionization happened rapidly, there should be a characteristic signature[2–4] visible against the smooth foreground in an all-sky spectrum. Here we report an all-sky spectrum between 100 and 200 MHz, corresponding to the redshift range $6 < z < 13$ for the 21-centimetre line. The data exclude a rapid reionization timescale of $\Delta z < 0.06$ at the 95% confidence level.

The observable differential brightness temperature caused by the redshifted 21-cm line from a volume of hydrogen gas in the intergalactic medium can be calculated from basic principles[5] and is:

$$\delta T_{21}(\boldsymbol{\theta}, z) \approx 27\ (1+\delta)\ x_{\mathrm{H\,I}} \left(1-\frac{T_\gamma}{T_\mathrm{S}}\right) \left(\frac{1+z}{10}\right)^{1/2} \mathrm{mK} \qquad (1)$$

where $\boldsymbol{\theta}$ is the position on the sky, z is the redshift of the gas, the factor of 27 mK comes from cosmological factors, δ is the local matter overdensity of the gas, $x_{\mathrm{H\,I}}$ is the neutral fraction of the gas, T_S is the 'spin' temperature that describes the relative population of the ground and excited states of the hyperfine transition, and T_γ is the temperature of the cosmic microwave background (CMB) radiation. The intensity of the 21-cm emission or absorption relative to the CMB has a strong dependence on the neutral fraction and the spin temperature, both of which are sensitive to the ultraviolet and X-ray radiation[2,3] from the formation of luminous sources, including early stars, galaxies and black holes.

During the reionization epoch, after the heating of the intergalactic medium, the spin temperature is expected to be much larger than the CMB temperature[6] ($T_\mathrm{S} \gg T_\gamma$) and the 21-cm perturbations will be seen in emission against the CMB and dominated by variations in the neutral fraction. Under the additional assumption that the local neutral fraction of the gas is not correlated to the local matter overdensity, the angle-averaged form of equation (1) can be reduced to:

$$\langle \delta T_{21}(\boldsymbol{\theta},z)\rangle_{\boldsymbol{\theta}} \equiv \delta \bar{T}_{21}(z) \approx 27 \left(\frac{1+z}{10}\right)^{1/2} \bar{x}_{H\,I}(z)\ \mathrm{mK} \qquad (2)$$

where we have explicitly written the redshift dependence of the mean neutral fraction as $\bar{x}_{H\,I}(z)$. This 'global' 21-cm signal should be observable in a measurement of the all-sky low-frequency radio spectrum through the mapping from redshift to frequency for the 21-cm line, according to: $\nu = 1{,}420/(1+z)$ MHz. Here, 1,420 MHz is the rest-frame frequency of the 21-cm line and the redshift range appropriate for the reionization epoch is $z > 6$. Fixing the overall amplitude factor in equation (2) through the choice of a particular cosmological model, for example, the WMAP7 best-fit ΛCDM model[7,8], the global 21-cm signal becomes a direct probe of the evolution of the mean neutral fraction of hydrogen gas in the intergalactic medium during the reionization epoch.

The global 21-cm signal is challenging to observe in practice because the low-frequency radio sky is dominated by intense synchrotron emission from our own Galaxy that is more than four orders of magnitude brighter than the signal. Galactic and extragalactic free-free emission provide additional foregrounds[7], as do numerous radio point sources from active galactic nuclei, radio galaxies and local Galactic objects. Radio-frequency interference from television, FM (frequency modulated) radio, low-Earth-orbit satellites, and other telecommunications transmitters are prolific and can be eight or ten orders of magnitude brighter than the astrophysical signal, even in geographically remote areas.

We deployed a custom-built, high-dynamic-range broadband radio spectrometer, called EDGES[9,10], at the Murchison Radio-astronomy Observatory in Western Australia, to measure the radio spectrum between 100 and 200 MHz. The instrument observed continuously for three months with low duty cycle and yielded the spectrum shown in Fig. 1—an average over nearly the entire southern celestial hemisphere.

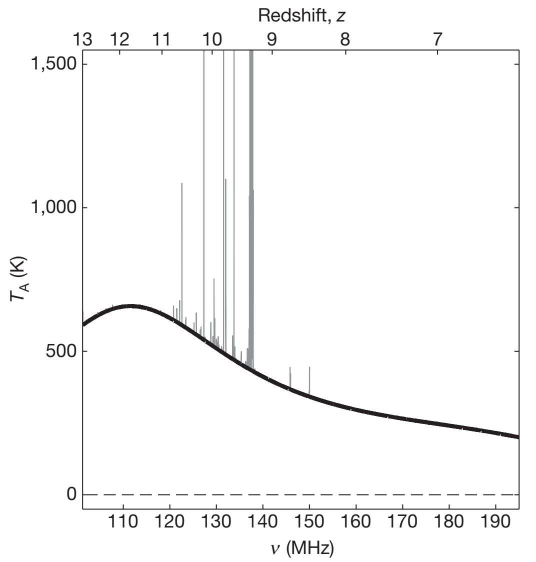


Figure 1 | Measured spectrum between 100 and 195 MHz. The spectrum corresponds to redshifts $13 > z > 6$. The grey spikes are spectral channels that experienced radio-frequency interference during the integration and are masked from the analysis. The shape and amplitude of the spectrum are dominated by Galactic synchrotron emission and modulated by the uncalibrated antenna bandpass, which causes the spectrum to roll off from the characteristic $T_\mathrm{F} \propto \nu^{-2.5}$ power-law form of the foregrounds at low and high frequencies. Any global 21-cm contribution in the spectrum is at the 20–30 mK level, approximately four orders of magnitude below the visible foreground emission. Thermal noise in the spectrum is 6 mK at 150 MHz using 1-MHz binned spectral resolution. The thermal noise increases at lower frequencies owing to the larger sky noise and lowered transmission efficiency of the antenna. Any 20-MHz sub-band in this spectrum can be fitted by a fifth-order polynomial, leaving residuals at or below the thermal noise level.

[1]Arizona State University, School of Earth and Space Exploration, Tempe, Arizona 85287, USA. [2]Massachusetts Institute of Technology, Haystack Observatory, Westford, Massachusetts 01886, USA.
*These authors contributed equally to this work.

To overcome the foreground signal, we relied on the expectation that all of the foregrounds have smooth continuum spectra that are well modelled as simple power laws in frequency[2,11] (or redshift), and that any uncertainties in the calibration of the spectrometer could introduce only smooth spectral deviations into the measurement. We designed the instrument to accomplish this goal by shortening the electrical path length between the antenna and the internal calibration source to less than a wavelength and by using an electrically compact dipole antenna. Additional details on the design of the instrument are presented in the Supplementary Information.

The observed spectrum was fitted by a model that consists of a 21-cm signal term and a polynomial term that accounts for the foregrounds and calibration uncertainties according to: $T_{\rm obs}(z) = \delta\bar{T}_{21}(z) + T_{\rm F}(z)$, where $T_{\rm F}(z) = \sum_{n=0}^{m} a_n z^n$ is the foreground term. The 21-cm term is given by equation (2) with

$$\tilde{x}_{\rm H\,I}(z) = \frac{1}{2}\left[\tanh\left(\frac{z - z_{\rm r}}{\Delta z}\right) + 1\right]$$

following the recent convention of the WMAP7 analysis. The free parameters in the 21-cm model are the reionization redshift $z_{\rm r}$ when the transition reaches 50% and the duration of reionization $\Delta z = (d\bar{x}_{\rm H\,I}/dz)^{-1}|_{\bar{x}_{\rm H\,I}=0.5}$. The model is sufficient to account for all of the visible features in the observed spectrum after spectral channels with radio-frequency interference have been masked from the data set.

We fitted the model to all available trials of $z_{\rm r}$ in the observed spectrum using an approximately 20-MHz subset of the spectrum centred on $\nu_{\rm r} = 1{,}420/(1 + z_{\rm r})$ MHz. In practice, we found that the order of polynomial that yields the best results is dependent on the trial redshift because the magnitude of the systematic structure in the spectrum varies with frequency. We used $m = 4$ for $z_{\rm r} < 9$ and $m = 5$ for $z_{\rm r} > 9$. By assuming that reionization was equally likely to have occurred at any redshift between $6 < z_{\rm r} < 13$ and treating each frequency trial as an independent measurement, the observations exclude reionization histories shorter than $\Delta z < 0.19$ at 68% statistical confidence. Systematic uncertainty is estimated through inspection of the distribution function of best-fit derivatives, Δz^{-1}, from all frequency trials. Under a null hypothesis, we expect the distribution to peak at $\Delta z^{-1} = 0$ with large deviations indicative of systematic errors. We set the systematic error as the 68th percentile of the derivative distribution, corresponding to $\Delta z_{\rm sys} = 0.21$. Combining statistical and systematic uncertainties in derivative space in quadrature yields our final confidence bounds of $\Delta z_{68} < 0.13$ at 68% combined confidence and $\Delta z_{95} < 0.06$ at 95% confidence. The excluded duration bounds as a function of reionization redshift $z_{\rm r}$ are plotted in Fig. 2. These constraints are sufficient to rule out the most rapid plausible reionization histories, although more general theoretical expectations[3] currently yield predictions of $1 < \Delta z < 10$. Our result extends findings by WMAP, which ruled out at a fiducial instantaneous transition at $z < 7$ and yielded a best-fit $z_{\rm r} = 10.5 \pm 1.2$.

The method demonstrated here is the only mechanism available at present to probe directly the derivative of the neutral fraction with redshift in the early Universe, and hence, to offer a unique way to constrain the reionization history. CMB anisotropy measurements probe reionization indirectly through an integral constraint on the optical depth to Thomson scattering of CMB photons and large-scale features in the E-mode polarization by free electrons in the intergalactic medium after reionization, whereas high-redshift quasar absorption spectra and Lyman-α galaxy surveys provide only a snapshot of the neutral fraction at the end of reionization.

Future enhancements to the EDGES instrument are forecast[11] to improve the constraints presented here by an order of magnitude and should be particularly valuable in combined analysis[12] with CMB and Lyman-α quasar spectra. The current measurement also serves as a pathfinder for extending the techniques to higher redshifts (lower frequencies) of $z \approx 20$ to search for absorption signatures[4,11,13] in the global 21-cm signal that should be more distinct than the reionization transition, but may be masked by larger foreground contributions. Such high-redshift global 21-cm observations may eventually provide unparalleled information about the ultraviolet emission and radiative feedback from the very first stars through the Wouthuysen–Field-effect-induced coupling of the kinetic and spin temperatures of the hydrogen gas in the intergalactic medium during the epoch of first light.

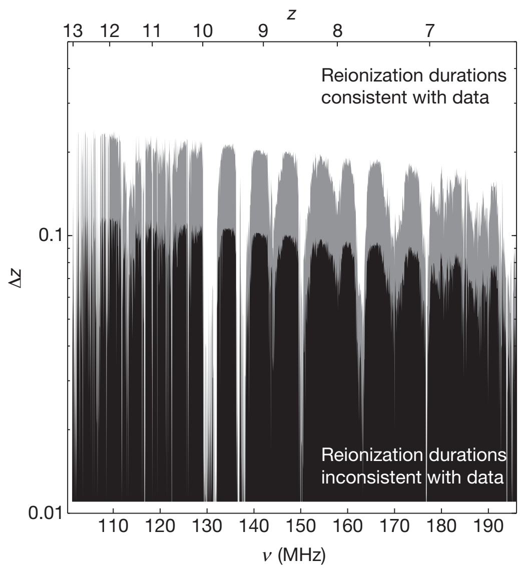


Figure 2 | Lower confidence bounds on the duration of the reionization transition. Statistical and systematic uncertainties are included. Grey indicates the 68% confidence bound and black the 95% bound. The white region is allowed by the data. The data rule out rapid reionization histories shorter than $\Delta z \approx 0.1$ for many redshifts between $6 < z < 13$. The two large gaps at redshifts $z \approx 9.5$ (138 MHz) and $z \approx 10$ (130 MHz) are at frequencies that require extensive radio-frequency interference excision because they fall into satellite and aircraft communication bands, respectively.

Received 25 May; accepted 20 October 2010.

1. Furlanetto, S. R., Oh, S. P. & Briggs, F. H. Cosmology at low frequencies: the 21 cm transition and the high-redshift Universe. *Phys. Rep.* **433,** 181–301 (2006).
2. Shaver, P. A., Windhorst, R. A., Madau, P. & de Bruyn, A. G. Can the reionization epoch be detected as a global signature in the cosmic background? *Astron. Astrophys.* **345,** 380–390 (1999).
3. Gnedin, N. Y. & Shaver, P. A. Redshifted 21 centimeter emission from the pre-reionization era. I. Mean signal and linear fluctuations. *Astrophys. J.* **608,** 611–621 (2004).
4. Furlanetto, S. R. The global 21-centimeter background from high redshifts. *Mon. Not. R. Astron. Soc.* **371,** 867–878 (2006).
5. Zaldarriaga, M., Furlanetto, S. R. & Hernquist, L. 21 centimeter fluctuations from cosmic gas at high redshifts. *Astrophys. J.* **608,** 622–635 (2004).
6. Madau, P., Meiksin, A. & Rees, M. J. 21 centimeter tomography of the intergalactic medium at high redshift. *Astrophys. J.* **475,** 429–444 (1997).
7. Larson, D. *et al.* Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: power spectra and WMAP-derived parameters. Preprint at 〈http://arxiv.org/abs/1001.4635〉 (2010).
8. Komatsu, E. *et al.* Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological interpretation. Preprint at 〈http://arxiv.org/abs/1001.4538〉 (2010).
9. Bowman, J. D., Rogers, A. E. E. & Hewitt, J. N. Toward empirical constraints on the global redshifted 21 cm brightness temperature during the epoch of reionization. *Astrophys. J.* **676,** 1–9 (2008).
10. Rogers, A. E. E. & Bowman, J. D. Spectral index of the diffuse radio background measured from 100 to 200 MHz. *Astron. J.* **136,** 641–648 (2008).
11. Pritchard, J. R. & Loeb, A. Constraining the unexplored period between the dark ages and reionization with observations of the global 21 cm signal. *Phys. Rev. D* **82,** 023006 (2010).
12. Pritchard, J. R., Loeb, A. & Wyithe, J. S. B. Constraining reionization using 21-cm observations in combination with CMB and Ly-α forest data. *Mon. Not. R. Astron. Soc.* **408** 57–70 (2010).
13. Pritchard, J. R. & Loeb, A. Evolution of the 21 cm signal throughout cosmic history. *Phys. Rev. D* **78,** 103511 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the NSF. J.D.B was supported during part of this work at the California Institute of Technology by NASA through a Hubble Fellowship awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy for NASA. This scientific work uses data obtained from the Murchison Radio-astronomy Observatory. We acknowledge the Wajarri Yamatji people as the traditional owners of the Observatory site. We thank CSIRO and Curtin University for logistical support and D. DeBoer, D. Herne, D. Emrich, M. Halleen and C. Halleen for on-site support.

Author Contributions Both authors contributed equally to the work in this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.D.B. (judd.bowman@asu.edu).

doi:10.1038/nature09539

Snapshots of cooperative atomic motions in the optical suppression of charge density waves

Maximilian Eichberger[1]*, Hanjo Schäfer[1]*, Marina Krumova[2], Markus Beyer[1], Jure Demsar[1,3], Helmuth Berger[4], Gustavo Moriena[5,6], Germán Sciaini[5,6]* & R. J. Dwayne Miller[5,6]

Macroscopic quantum phenomena such as high-temperature superconductivity, colossal magnetoresistance, ferrimagnetism and ferromagnetism arise from a delicate balance of different interactions among electrons, phonons and spins on the nanoscale[1]. The study of the interplay among these various degrees of freedom in strongly coupled electron–lattice systems is thus crucial to their understanding and for optimizing their properties. Charge-density-wave (CDW) materials[2], with their inherent modulation of the electron density and associated periodic lattice distortion, represent ideal model systems for the study of such highly cooperative phenomena. With femtosecond time-resolved techniques, it is possible to observe these interactions directly by abruptly perturbing the electronic distribution while keeping track of energy relaxation pathways and coupling strengths among the different subsystems[3–7]. Numerous time-resolved experiments have been performed on CDWs[8–13], probing the dynamics of the electronic subsystem. However, the dynamics of the periodic lattice distortion have been only indirectly inferred[14]. Here we provide direct atomic-level information on the structural dynamics by using femtosecond electron diffraction[15] to study the quasi two-dimensional CDW system 1T-TaS_2. Effectively, we have directly observed the atomic motions that result from the optically induced change in the electronic spatial distribution. The periodic lattice distortion, which has an amplitude of ~0.1 Å, is suppressed by about 20% on a timescale (~250 femtoseconds) comparable to half the period of the corresponding collective mode. These highly cooperative, electronically driven atomic motions are accompanied by a rapid electron–phonon energy transfer (~350 femtoseconds) and are followed by fast recovery of the CDW (~4 picoseconds). The degree of cooperativity in the observed structural dynamics is remarkable and illustrates the importance of obtaining atomic-level perspectives of the processes directing the physics of strongly correlated systems.

1T-TaS_2 is one of the most-studied quasi-two-dimensional CDW systems[16–19]. It has a simple crystalline structure, consisting of planes of hexagonally arranged tantalum (Ta) atoms, sandwiched by two sulphur (S) layers coordinating the central Ta atom in an octahedral arrangement[16,17]. In the low-temperature CDW phase, the conduction electron density becomes modulated, modifying the forces among the ions and generating a periodic lattice distortion (PLD) with a periodicity of ~12 Å. This effect is illustrated in Fig. 1a, b together with the corresponding potential energy surfaces, $U(Q)$, where Q is the generalized coordinate of the atomic displacements. The corresponding changes in $U(Q)$ result in a shift of the equilibrium atomic positions to introduce a PLD. In 1T-TaS_2, the transition from its metallic, unmodulated, phase to an incommensurate CDW phase (ICP) happens at 550 K. At 350 K, a transition to a nearly commensurate CDW phase (NCCP) occurs, where the amplitude of the PLD increases abruptly from 0.03 to 0.1 Å and the CDW wavevector undergoes a sudden angular rotation from $\phi = 0°$ to $\sim 12.3°$ with respect to the fundamental lattice vector of the host (unreconstructed) lattice. Finally, a transition to a commensurate CDW phase (CCP) takes place at 180 K, with $\phi = 13.9°$ and a $\sqrt{13} \times \sqrt{13}$ periodicity[17]. This phase transition is characterized by the appearance of the gap throughout the Fermi surface, and is argued to be due to Mott localization[20]. The appearance of the ICP can be described by the standard Peierls model[2]. According to this, in low-dimensional systems the divergence in the static electronic susceptibility at the wavevector $2\mathbf{k}_F$, connecting parallel Fermi surfaces at $\mathbf{k}_F$ and $-\mathbf{k}_F$ (where $\mathbf{k}_F$ is the Fermi wavevector), gives rise to an instability of conduction electrons against the formation of the electron density modulation. Indeed, the comparison of the topology of the Fermi surface with parallel sections that can be connected by the modulation wavevector favours the standard Peierls model for the emergence of the ICP[16–18]. The nature of the CCP and

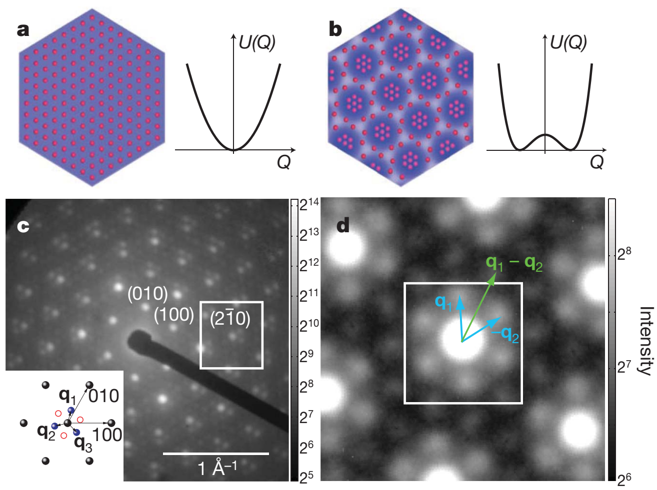


Figure 1 | FED data in the NCCP in 1T-TaS_2. **a**, **b**, Schematic real-space images of Ta atoms and the electron density in the metallic (**a**) and the CDW (**b**) states together with the corresponding potential energy, U, as a function of generalized coordinate, Q. **c**, The diffraction pattern of 1T-TaS_2 at 200 K (intensity is shown on a logarithmic scale in arbitrary units). Each Bragg reflection is surrounded by six first-order CDW reflections at the scattering wave-vectors $\mathbf{q}_i$, which each has an out-of-plane component of $\pm 1/3c^*$ (red and blue circles, respectively, in inset). The projections of the $\mathbf{q}_i$ on the basal plane, with a modulus of $\sim 0.28a^*$, are tilted away from the closest fundamental lattice vector by an angle $\phi \approx 12.3°$. **d**, Magnified view of the diffraction intensity (I) near the (2$\bar{1}$0) Bragg peak (see box in **c**; for presentation purposes the diffraction image was symmetrized with respect to the six-fold axis). The secondary CDW reflections at the wavevector corresponding to the difference of the wavevectors of the first-order CDW peaks[20] are clearly resolved.

[1]Physics Department and Center of Applied Photonics and Zukunftskolleg, University of Konstanz, D-78457 Konstanz, Germany. [2]Department of Chemistry, University of Konstanz, D-78457 Konstanz, Germany. [3]Complex Matter Department, Jozef Stefan Institute, SI-1000 Ljubljana, Slovenia. [4]Physics Department, EPFL CH-1015 Lausanne, Switzerland. [5]Institute for Optical Sciences and Departments of Chemistry and Physics, University of Toronto, Toronto, Ontario M5S 3H6, Canada. [6]Max Planck Research Department for Structural Dynamics, Department of Physics, University of Hamburg, Centre for Free Electron Laser Science, DESY, D-22607 Hamburg, Germany.
*These authors contributed equally to this work.

the NCCP, as well as of the ICP–NCCP and NCCP–CCP transitions, is, however, still under debate[18]. Recently, 1T-TaS_2 received additional attention owing to the observation there of superconductivity below 5 K under high pressure[19].

In this study, we investigated the structural dynamics of the PLD in 30-nm-thick, free-standing slices of 1T-TaS_2. We performed femtosecond electron diffraction (FED) experiments in transmission geometry along the *c* axis, that is, perpendicular to the TaS_2 layers. The films were photoexcited with 140-fs optical pulses, and 50-keV electrons, in bursts of ≤250 fs, were used to monitor the structural changes by recording time-delayed diffraction patterns. The diffraction pattern of the NCCP (200 K) recorded in this set-up is shown in Fig. 1c together with the assignment of some of the scattering vectors. The intense peaks are the Bragg reflections of the host lattice. Each of the Bragg peaks is surrounded by six weak satellite peaks originating from the PLD, with modulation wavevectors $\mathbf{q}_i$ (ref. 21), illustrated in Fig. 1c (inset) and Fig. 1d.

Figure 2a–e shows the time evolution of the relative change of the diffraction signal in the vicinity of a Bragg peak, following photoexcitation. The corresponding traces of the relative changes in the Bragg peaks ($\Delta I_{Bragg}/I_{Bragg}$), the inelastic background ($\Delta I_{bckg}/I_{bckg}$) and the CDW peaks ($\Delta I_{CDW}/I_{CDW}$) are shown in Fig. 2f (see also Supplementary Fig. 3). The intensity of the CDW peaks (the satellites of the Bragg peaks), I_{CDW}, is suppressed by ~30% on the timescale of hundreds of femtoseconds. The corresponding suppression of the PLD gives rise to more-efficient scattering into the Bragg reflections of the host lattice, manifested by an increase of the Bragg peak intensity, I_{Bragg}, by ~15%. In the CDW state, the presence of the PLD suppresses I_{Bragg} similarly to the effect of thermally induced disorder; that is, the presence of PLD can be looked upon as an effective Debye–Waller effect. The decrease in I_{CDW} and the accompanying increase in I_{Bragg} thus illustrate a cooperative phenomenon in which the optically induced redistribution of electron density efficiently decreases the PLD amplitude. Because I_{CDW} is proportional to the square of the atomic displacements[22], the resulting suppression of I_{CDW}, by ~30%, corresponds to ~16% change in atomic displacements (~0.02 Å). Following the initial increase, I_{Bragg} is found to partially recover on the 350-fs timescale. This recovery is accompanied by an increase in the inelastic background intensity, I_{bckg}—see the intensity changes in the area indicated by the circle in Fig. 2e for the frames between 300 (Fig. 2c) and 5,800 fs (Fig. 2e). This process can be attributed to the generation of phonons with non-zero momentum ($q \neq 0$); hence, I_{Bragg} is reduced owing to the conventional Debye–Waller effect, leading to an increase in the inelastic background.

A noteworthy feature of the data shown in Fig. 2f, and elaborated on in Fig. 3a, is the apparent difference between the dynamics of $\Delta I_{Bragg}/I_{Bragg}$ and $\Delta I_{CDW}/I_{CDW}$. Although the maximum in I_{Bragg} is reached at a time delay of ~300 fs (Fig. 3a, dashed vertical line), the minimum in I_{CDW} is reached at a time delay of ~500 fs (Fig. 3a, solid vertical line), at which point I_{Bragg} has decreased from its maximum. This difference can be naturally explained by considering the effects of both the suppression of the PLD and the increase in the $q \neq 0$ phonon population on the two diffraction intensities. For the case of I_{Bragg}, the first effect gives rise to its increase as the periodicity of the host lattice is enhanced, and the increase in the $q \neq 0$ phonon population (the Debye–Waller effect) has the opposite effect. Indeed, from the fast recovery of I_{Bragg} and the corresponding increase in I_{bckg} it follows that the energy transfer from electrons to $q \neq 0$ phonons in 1T-TaS_2 takes place on the timescale of a few hundred femtoseconds ($\tau_{e-ph} \approx 350$ fs). For I_{CDW}, both the displacive excitation of highly correlated atomic motions and phonon-induced disorder contribute to its suppression, explaining the longer timescale on which the minimum of I_{CDW} is reached.

It is instructive to compare the structural dynamics data with those of the electronic subsystem. We have performed all-optical pump–probe measurements, where the dynamics are mainly sensitive to the changes in the electronic properties. The photoinduced reflectivity change (Fig. 2g) shows a rapid onset on the 100-fs timescale, followed by a fast recovery with a decay time of 150 fs and subsequent slower decay with a relaxation time of ~4 ps, which is nearly identical to the

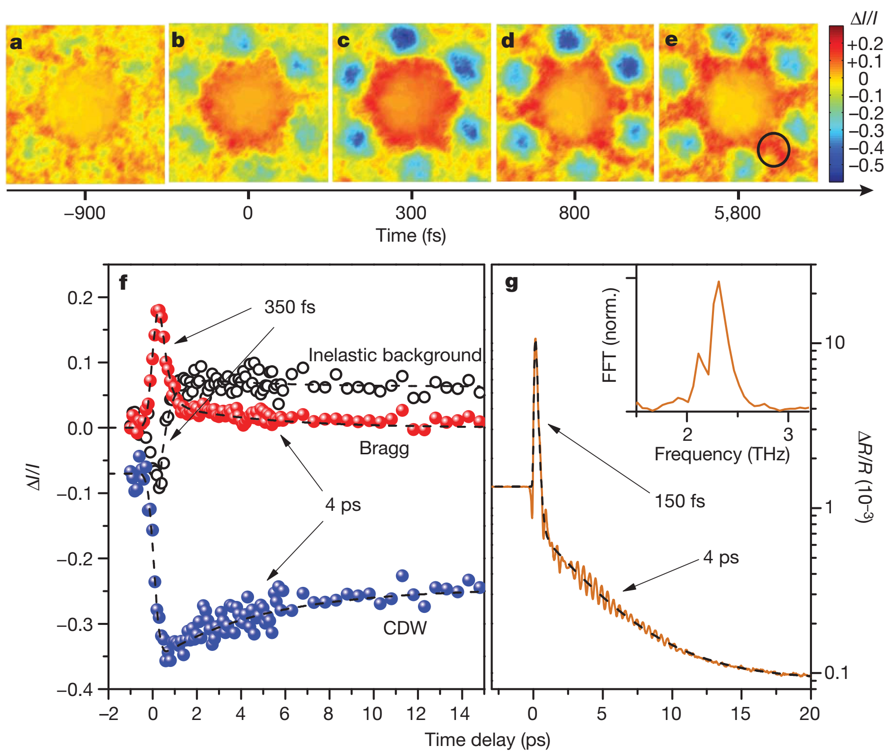


Figure 2 | Time evolution of the diffraction intensities following photoexcitation with a fluence of 2.4 mJ cm^{-2}. **a**–**e**, Evolution of Bragg, CDW and inelastic background intensities illustrated as relative change in the diffraction pattern at several time delays following photoexcitation with a fluence of 2.4 mJ cm^{-2} and a photon energy of 3.2 eV (see also Supplementary Fig. 4 and Supplementary Information). These images were obtained by averaging (area enclosed by the box in Fig. 1d) over all individual Bragg reflections to increase the signal-to-noise ratio. The circle in **e** represents the area over which the inelastic background intensity was monitored. **f**, Corresponding dynamics of $\Delta I_{Bragg}/I_{Bragg}$, $\Delta I_{CDW}/I_{CDW}$ and $\Delta I_{bckg}/I_{bckg}$ with fits to the data (dashed lines) and the extracted timescales. The suppression of the PLD, that is, the CDW peak intensity at negative time delays, is due to an increase in the sample temperature caused by the photoexcitation pulse train (accumulative heating). The initial drop in I_{bckg} is an artefact, a result of the decrease in the diffraction intensity of the nearby CDW peaks, whose tails extend well into the region where the inelastic background was evaluated (**e**). **g**, Dynamics of the differential reflectivity change, $\Delta R/R$, at 1.55 eV (800 nm), recorded at the same initial temperature and the same excitation energy density, together with the fit (dashed line). The signal has been offset vertically for presentation purposes. The oscillatory response corresponds to the coherently excited amplitude mode at 2.3 THz and phonon mode at 2.1 THz (refs 9, 11, 21). Inset, fast Fourier transform (FFT) of the oscillatory component.

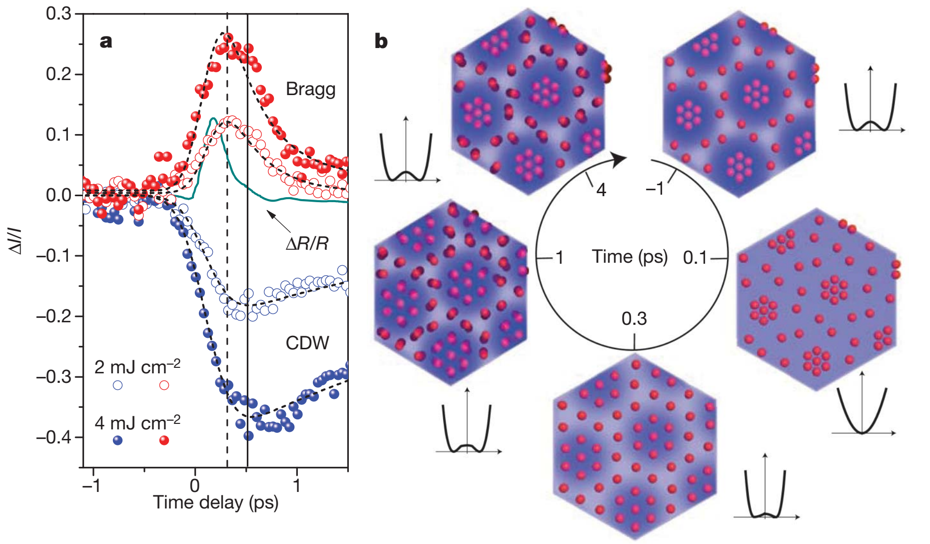


Figure 3 | Early time dynamics and emerging time evolution of the CDW state in 1T-TaS_2 on photoexcitation. **a**, Data were recorded with a 40-fs time step at two excitation fluences, and compared with the optical $\Delta R/R$ data. The maximum induced changes in the Bragg (dashed vertical line) and CDW (solid vertical line) peaks were achieved ~300 fs and, respectively, ~500 fs after photoexcitation. **b**, The evolution of the real-space structure of the Ta plane of 1T-TaS_2 following photoexcitation with an intense optical pulse (circles represent Ta atoms and blue shading represents the density of conduction electrons; the amplitudes are strongly exaggerated). Before photoexcitation ($t \approx -1$ ps), the Ta atoms are periodically displaced from their pure 1T structure, forming a nearly commensurate CDW. Intense perturbation of the electronic system gives rise to smearing of the electron density modulation ($t \approx 0.1$ ps), driving the lattice towards the undistorted state (at $t \approx 0.3$ ps, the hexagonal symmetry of the pure 1T phase is nearly recovered). In parallel, the energy is transferred from the electronic subsystem to phonons on the 300-fs timescale, resulting in recovery of the electron density modulation and thermal disordering of the lattice ($t \approx 1$ ps). The CDW order is recovered at $t \approx 4$ ps, after which time the sample is thermalized at a somewhat higher temperature.

CDW recovery time observed in the FED studies. The short decay timescale is identical to the one obtained in the NCCP by time- and angle-resolved photoemission spectroscopy[11] (tr-ARPES) and can be attributed to the electron–phonon energy transfer. Because the electron–phonon scattering rate is strongly momentum dependent ($\propto 1/q$), it is quite natural to observe shorter time constants in optics and tr-ARPES than in FED. In the former experiments the signal is dominated by the energy transfer to $q \approx 0$ phonons, whereas in the latter the behaviour of I_{Bragg} and I_{bckg} is governed by the population of large-q phonons. In the optical data, owing to their high signal-to-noise ratio, in addition to the electronic response a weak oscillatory signal is observed. The main mode observed at 2.3 THz is the totally symmetric amplitude mode[9,11,23] of the CDW, whose amplitude is apparently smaller than the noise level in the FED data.

Despite the fact that a large amount of energy is transferred to phonons on the subpicosecond timescale, the system is not yet in thermal equilibrium 1 ps after photoexcitation. The recovery of the PLD amplitude is clearly observed in I_{Bragg} and I_{CDW}. This timescale is well decoupled from both subpicosecond timescales. By fitting the recovery of I_{CDW} with an exponential decay, we obtain a CDW recovery time of $\tau_{rec} \approx 4$ ps. As this timescale is much longer than the oscillation period of the amplitude mode, it is reasonable to assume that the electronic part of the order parameter follows the PLD on the aforementioned timescale. Here the process that governs the CDW recovery dynamics is the thermalization with the longer-wavelength acoustic phonons by means of anharmonic phonon decay. Indeed, the characteristic linewidths of the low-energy optical phonons[23] are about 10 cm^{-1}, corresponding to lifetimes of 3 ps. The two distinct relaxation timescales, one of the order of 100-fs and the other of several picoseconds, are commonly observed in optical experiments in CDWs[8,10]. From the direct structural dynamics and optical data on 1T-TaS_2, we can conclude that the longer timescale describes the recovery of the coupled electron–lattice order parameter and that the shorter timescale corresponds to the partial recovery of the electronic part alone[14].

To determine the time constant for the electronic suppression of the PLD, which leads to an increase in I_{Bragg}, we analysed its dynamics. By fitting (Supplementary Fig. 5 and Supplementary Information) the I_{Bragg} trace, taking into account the finite optical and electron pulse widths, we determined a timescale of $\tau_{supp} \approx 250 \pm 70$ fs for the PLD suppression.

Information complementary to the above findings comes from considering the energy flow following photoexcitation. In the experiments with fluences $F = 2$–4 mJ cm^{-2}, no signature of the NCCP–ICP transition is observed. Only at $F = 4.8$ mJ cm^{-2} is the photoinduced NCCP–ICP transition realized (Supplementary Fig. 6 and Supplementary Information), as demonstrated by a strong suppression of the CDW peak intensity and a rotation of the primary CDW wavevectors, $\mathbf{q}_i$, by $\phi \approx 10°$. Using the literature values of the optical constants and the overall specific heat (Supplementary Information), we obtained a temperature increase of ~180 K at $F = 4.8$ mJ cm^{-2}. This implies that the energy needed to drive the phase transition is comparable to the energy required simply to heat the sample across the phase transition. The rapid energy transfer from the electronic system to phonons ($\tau_{e\text{–}ph} \approx 150$–$350$ fs), which is competing with the electronically driven PLD suppression process ($\tau_{supp} \approx 250$ fs), and the fact that the electronically excited symmetric amplitude mode does not map into the rotation of the CDW wavevector, suggest that the NCCP–ICP transition can be driven only thermally.

The direct structural information obtained with FED, supported by time-resolved optical and published tr-ARPES[11] data, enabled us to elucidate the dynamics of the coupled electron–lattice order parameter (Fig. 3b). Strong photoexcitation and subsequent electron–electron scattering creates a high density of electron–hole pairs within ≤100 fs, raising the effective electronic temperature to several thousand kelvin. The electronic modulation is thereby strongly suppressed, modifying the potential energy surface $U(Q)$. The collapse of the double-well potential brings about highly cooperative atomic motions towards a new quasi-equilibrium. This coherent process is, however, accompanied by the rapid recovery of the double-well potential due to cooling of the electronic subsystem through the electron–phonon scattering, which also takes place on the subpicosecond timescale[8–13]. The resulting suppression of the PLD amplitude, by ~20% (0.02 Å), happens within a time of $\tau_{supp} \approx 250 \pm 70$ fs, that is, about half the period of the amplitude mode, ~440 fs (refs 9, 11). After a time delay of $t \approx 300$ fs, the periodicity of the underlying lattice has increased and the amplitude of the PLD has decreased. By $t \approx 1$ ps, the electronic modulation has been largely recovered and the electrons have transferred the energy to $q \neq 0$ phonons, randomizing the atomic motions. Finally, the coupled electron–lattice order parameter recovers on the timescale of ~4 ps, when the excess energy is

redistributed by further thermalization with low-energy acoustic phonons.

These results demonstrate the extreme robustness of the PLD in 1T-TaS_2 against electronic excitation triggered by a femtosecond optical pulse. By contrast, in the insulating CCP the gap is fully suppressed[11] at only one-tenth of the absorbed energy density used in our study. The large difference in the two energy densities presents a strong argument that the NCCP–CCP transition is indeed Mott driven[11,19].

The present work illustrates the importance of directly observing atomic motions on timescales short enough to follow even the effect of non-equilibrium electronic distributions on strongly correlated lattice dynamics. In this respect, the introduction of table-top FED systems[24–26] with sufficient brightness and time resolution[24,26] is opening new pathways to the investigation of a myriad of cooperative systems in which electron–lattice correlations have an important role[26–30]. In systems with reduced dimensionality, such as quasi-one-dimensional and quasi-two-dimensional systems, in which structural changes have a predominantly in-plane character, the use of FED may be particularly advantageous. Because information about structural dynamics over the entire two-dimensional Brillouin zone is obtained in a single experimental run by FED, it is easy to distinguish between different processes that give rise to changes in the diffraction intensities, as in the case of 1T-TaS_2. Moreover, with further instrumental improvements, for example an increase in the signal-to-noise ratio, FED could be used to find signatures of lattice modulations, which may be difficult to determine by means of static diffraction methods, much like modulation optical spectroscopy is used to determine the electronic band structure in solids.

METHODS SUMMARY

In the present study, we used electron bunches of $\leq$250-fs duration containing 4,000 electrons, each of which had a kinetic energy of 50 keV. The electron beam (spot size, 150 μm) was collimated by a magnetic lens to scatter from the sample and generate a diffraction pattern downstream. The diffraction patterns formed on a phosphor screen after being intensified by a multichannel plate, and were recorded using a charge-coupled-device camera. The background pressures were 10^{-9} and 10^{-7} mbar in the electron gun and the sample chamber sections, respectively. We made the measurements in transmission mode at a repetition rate of 1 kHz. In this geometrical configuration, spatiotemporal mismatch and surface charging effects are negligible. The electron pulse duration was characterized using a recently developed electron/laser-pulse cross-correlation method based on ponderomotive scattering and *N*-body simulations. Photoinduced structural changes were initiated by 387-nm, 140-fs pump pulses focused to a spot with a full-width at half-maximum of 350 μm. The overall instrumental response time was 240–290 fs. The sample temperature, of 200 K, was achieved by using a cold finger attached to a well-conducting sample holder made of oxygen-free copper. We measured the temperature *in situ* using a calibrated temperature sensor. The 30-nm-thick, single-crystalline 1T-TaS_2 slices, ~200 μm × 200 μm in size (Supplementary Figs 2 and 3), were obtained by cleaving a thicker single crystal using an ultramicrotome. The slices were picked up from the water surface using a host copper mesh. All-optical measurements were performed in reflection geometry using 60-fs optical pulses (carrier wavelength, 800 nm) at a repetition rate of 100 kHz. FED and all-optical experiments were carried out with the same excitation energy density and under the same sample temperature conditions.

Received 17 May; accepted 24 September 2010.
Published online 24 November 2010.

1. Imada, M., Fujimori, A. & Tokura, Y. Metal-insulator transitions. *Rev. Mod. Phys.* **70,** 1039–1263 (1998).
2. Grüner, G. *Density Waves in Solids* (Addison-Wesley, 1994).
3. Kusar, P. *et al.* Controlled vaporization of the superconducting condensate in cuprate superconductors by femtosecond photoexcitation. *Phys. Rev. Lett.* **101,** 227001 (2008).
4. Ogasawara, T. *et al.* General features of photoinduced spin dynamics in ferromagnetic and ferrimagnetic compounds. *Phys. Rev. Lett.* **94,** 087202 (2005).
5. Averitt, R. D. *et al.* Ultrafast conductivity dynamics in colossal magnetoresistance manganites. *Phys. Rev. Lett.* **87,** 017401 (2001).
6. Rini, M. *et al.* Control of the electronic phase of a manganite by mode-selective vibrational excitation. *Nature* **449,** 72–74 (2007).
7. Kübler, C. *et al.* Coherent structural dynamics and electronic correlations during an ultrafast insulator-to-metal phase transition in VO_2. *Phys. Rev. Lett.* **99,** 116401 (2007).
8. Demsar, J., Biljakovic, K. & Mihailovic, D. Single particle and collective excitations in the one-dimensional charge density wave solid $K_{0.3}MoO_3$ probed in real time by femtosecond spectroscopy. *Phys. Rev. Lett.* **83,** 800–803 (1999).
9. Demsar, J. *et al.* Femtosecond snapshots of gap-forming charge-density-wave correlations in quasi-two-dimensional dichalcogenides 1T-TaS_2 and 2H-$TaSe_2$. *Phys. Rev. B* **66,** 041101 (2002).
10. Yusupov, R. V. *et al.* Single-particle and collective mode couplings associated with 1- and 2-directional electronic ordering in metallic RTe_3 (R = Ho, Dy, Tb). *Phys. Rev. Lett.* **101,** 246402 (2008).
11. Perfetti, L. *et al.* Time evolution of the electronic structure of 1T-TaS_2 through the insulator-metal transition. *Phys. Rev. Lett.* **97,** 067402 (2006).
12. Schmitt, F. *et al.* Transient electronic structure and melting of a charge density wave in $TbTe_3$. *Science* **321,** 1649–1652 (2008).
13. Tomeljak, A. *et al.* Dynamics of photoinduced charge-density-wave to metal phase transition in $K_{0.3}MoO_3$. *Phys. Rev. Lett.* **102,** 066404 (2009).
14. Schäfer, H. *et al.* Disentanglement of the electronic and lattice parts of the order parameter in a 1D charge density wave system probed by femtosecond spectroscopy. *Phys. Rev. Lett.* **105,** 066402 (2010).
15. Miller, R. J. D. *et al.* 'Making the molecular movie': first frames. *Acta Crystallogr. A* **66,** 137–156 (2010).
16. Williams, P. M., Parry, G. S. & Scruby, C. B. Diffraction evidence for Kohn anomaly in 1T-TaS_2. *Phil. Mag.* **29,** 695–699 (1974).
17. Wilson, J. A. Di Salvo, F. J. & Mahajan, S. Charge-density waves and superlattices in the metallic layered transition-metal dichalcogenides. *Adv. Phys.* **24,** 117–201 (1975).
18. Clerc, F. *et al.* Lattice-distortion-enhanced electron-phonon coupling and Fermi surface nesting in 1T-TaS_2. *Phys. Rev. B* **74,** 155114 (2006).
19. Sipos, B. *et al.* From Mott state to superconductivity in 1T-TaS_2. *Nature Mater.* **7,** 960–965 (2008).
20. Fazekas, P. & Tosatti, E. Electrical, structural and magnetic-properties of pure and doped 1T-TaS_2. *Phil. Mag. B* **39,** 229–244 (1979).
21. Scruby, C. B., Williams, P. M. & Parry, G. S. The role of charge density waves in structural transformations of 1T-TaS_2. *Phil. Mag.* **31,** 255–274 (1975).
22. Als-Nielsen, J. & McMorrow, D. *Elements of Modern X-ray Physics* Ch. 4.4.5 (Wiley, 2001).
23. Duffey, J. R., Kirby, R. D. & Coleman, R. V. Raman scattering from 1T-TaS_2. *Solid State Commun.* **20,** 617–621 (1976).
24. Siwick, B. J., Dwyer, J. R., Jordan, R. E. & Miller, R. J. D. An atomic-level view of melting using femtosecond electron diffraction. *Science* **302,** 1382–1385 (2003).
25. Chergui, M. & Zewail, A. H. Electron and X-ray methods of ultrafast structural dynamics: advances and applications. *ChemPhysChem* **10,** 28–43 (2009).
26. Sciaini, G. *et al.* Electronic acceleration of atomic motions and disordering in bismuth. *Nature* **458,** 56–59 (2009).
27. Sokolowski-Tinten, K. *et al.* Femtosecond X-ray measurement of coherent lattice vibrations near the Lindemann stability limit. *Nature* **422,** 287–289 (2003).
28. Johnson, S. L. *et al.* Directly observing squeezed phonon states with femtosecond X-ray diffraction. *Phys. Rev. Lett.* **102,** 175503 (2009).
29. Fritz, D. M. *et al.* Ultrafast bond softening in bismuth: mapping a solid's interatomic potential with X-rays. *Science* **315,** 633–636 (2007).
30. Beaud, P. *et al.* Ultrafast structural phase transition driven by photoinduced melting of charge and orbital order. *Phys. Rev. Lett.* **103,** 155702 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to acknowledge discussions with V. V. Kabanov, T. Dekorsy, D. Mihailovic, U. Bovensiepen and M. Wolf, and thank A. Nagy for help in preparing the video in Supplementary Material. This research was supported by the Sofja Kovalevskaja Award of the Alexander von Humboldt Foundation, the Center for Applied Photonics and Zukunftskolleg at the University of Konstanz, the Natural Science and Engineering Research Council of Canada and the Canada Foundation for Innovation. M.E. acknowledges financial support through the Stiftung der Deutschen Wirtschaft. H.B. acknowledges financial support from the Swiss NSF and the NCCR MaNEP.

Author Contributions R.J.D.M. and J.D. directed this work. H.B. grew 1T-TaS_2 single crystals. M.E. and M.K. prepared thin films and performed transmission electron microscopy characterization. G.S., G.M., M.E. and H.S. performed the FED experiments at the University of Toronto. M.B., M.E. and H.S. performed the optical pump–probe experiments at the University of Konstanz. M.E., H.S. and J.D. performed the data analysis. J.D., G.S. and R.J.D.M. wrote the paper. All authors contributed to discussions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.J.D.M. (dmiller@lphys.chem.utoronto.ca) or J.D. (jure.demsar@uni-konstanz.de).

doi:10.1038/nature09618

Ice-sheet acceleration driven by melt supply variability

Christian Schoof[1]

Increased ice velocities in Greenland[1] are contributing significantly to eustatic sea level rise. Faster ice flow has been associated with ice–ocean interactions in water-terminating outlet glaciers[2] and with increased surface meltwater supply to the ice-sheet bed inland. Observed correlations between surface melt and ice acceleration[2–6] have raised the possibility of a positive feedback in which surface melting and accelerated dynamic thinning reinforce one another[7], suggesting that overall warming could lead to accelerated mass loss. Here I show that it is not simply mean surface melt[4] but an increase in water input variability[8] that drives faster ice flow. Glacier sliding responds to melt indirectly through changes in basal water pressure[9–11], with observations showing that water under glaciers drains through channels at low pressure or through interconnected cavities at high pressure[12–15]. Using a model that captures the dynamic switching[12] between channel and cavity drainage modes, I show that channelization and glacier deceleration rather than acceleration occur above a critical rate of water flow. Higher rates of steady water supply can therefore suppress rather than enhance dynamic thinning[16], indicating that the melt/dynamic thinning feedback is not universally operational. Short-term increases in water input are, however, accommodated by the drainage system through temporary spikes in water pressure. It is these spikes that lead to ice acceleration, which is therefore driven by strong diurnal melt cycles[4,14] and an increase in rain and surface lake drainage events[8,17,18] rather than an increase in mean melt supply[3,4].

The effective pressure in the subglacial drainage system, defined as overburden minus basal water pressure, controls coupling between ice and bed: lower effective pressure weakens the ice–bed contact and permits faster sliding[9–11]. Effective pressure is controlled by subglacial drainage, which occurs through two principal types of conduit (Fig. 1): Röthlisberger channels[19,20] are kept open by a balance between a widening of the channel by wall melting due to heat dissipation in the water flow, and a narrowing that results from the inward creeping motion of the surrounding ice. By contrast, cavities[11,21,22] are formed where ice is forced upwards by horizontal sliding over protrusions on the glacier bed. This opens a gap in the lee of the protrusion, with gap size controlled by the opening rate due to sliding and by creep closure of the cavity roof.

An increase in effective pressure leads to faster creep closure. In an equilibrium channel, this must be balanced by greater wall melt. Greater wall melt in turn requires higher discharge and, thus, a larger channel. Röthlisberger channels therefore increase in size with increasing effective pressure (decreasing water pressure). This causes water flow from smaller channels into larger ones, favouring the formation of an arterial network with few main channels at low water pressure[19,23]. Cavities differ from channels as their size is not controlled by wall melt and increases rather than decreases with water pressure. A reduction in effective pressure suppresses creep closure and allows larger cavities to form[11,22]. This favours macroporous behaviour[24] with spatially distributed drainage along the ice–bed interface and water discharge increasing with water pressure. The abundance of channels relative to cavities therefore determines whether water pressure is low or high in the steady state: channels can efficiently transport water at high effective pressure whereas cavities require low effective pressure to transport the same flux. Past models[23,25], however, do not capture switches from cavities to channels in spatially extended drainage or the formation of an arterial network, and cannot predict the spatial configuration of the drainage system.

Here I unify the description of cavities and channels and predict how spatially extended drainage systems can switch from cavities to channels and back. The basic physics of cavities and channels can be captured in a single equation for the cross-sectional area, S, of a subglacial conduit, which can be a channel or cavity (Supplementary Information and Fig. 1):

$$\frac{\mathrm{d}S}{\mathrm{d}t} = c_1 Q \Psi + u_b h - c_2 N^n S \qquad (1)$$

where Q is the water discharge, Ψ is the hydraulic gradient along the conduit and $N = p_i - p_w$ is the effective pressure in the conduit (ice

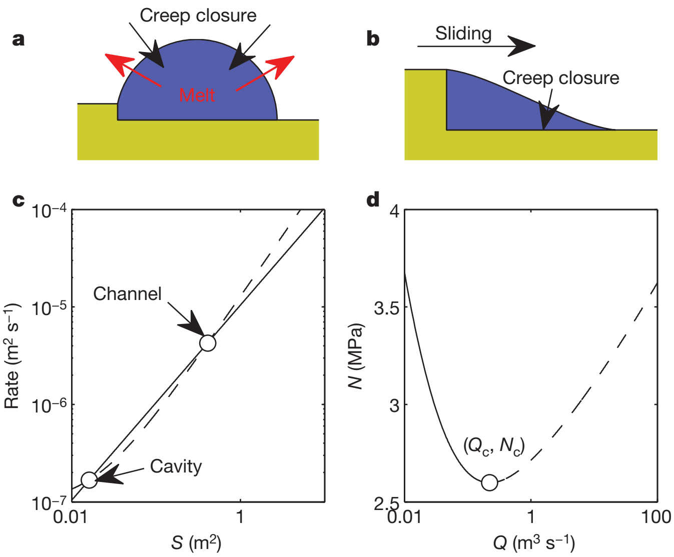


Figure 1 | Properties of a single conduit. **a**, **b**, Physics of channels (**a**) and cavities (**b**). **c**, Conduit opening rate, $c_1 Q\Psi + u_b h$ (dashed line), and closure rate, $c_2 N^n S$ (solid line), plotted against S. **d**, Steady-state N versus Q in a conduit (equation (2)). Parameter values are given in Methods Summary. Each conduit can generally attain one of two equilibria (points of intersection given as circles in **c**). These can be identified as channel and cavity. The larger (channel) equilibrium is prone to instability[20]: if perturbed to slightly larger size, the conduit will continue to grow (opening rate exceeds closing rate to the right of the intersection). In a network of conduits, this eventually leads to one channel growing at the expense of all other nearby ones. The cavity equilibrium, by contrast, is stable, and cavities of similar size can coexist. In the steady state, effective pressure increases with discharge in a channel (increasing N makes the closure curve steeper, moving the channel intersection in **c** to larger values of S), and decreases with discharge in a cavity. A conduit becomes a channel above a critical discharge, Q_c (dashed curve in **d**), and remains a cavity below Q_c.

[1]Department of Earth and Ocean Sciences, University of British Columbia, 6339 Stores Road, Vancouver, British Columbia V6T 1Z4, Canada.

overburden, p_i, minus water pressure, p_w). Q is related to S and Ψ through the Darcy–Weisbach law[26], $Q = c_3 S^{\alpha} |\Psi|^{-1/2} \Psi$, where $\alpha = 5/4$ and c_3 is related to the Darcy–Weisbach friction factor. The first term in equation (1) is the rate of conduit opening due to wall melting, the second is the rate of opening due to sliding of ice at speed u_b over bed protrusions of size h and the third is conduit roof closure due to viscous creep; c_1, c_2 and n are constants related to the latent heat of fusion and ice viscosity.

In the steady state, the effective pressure and discharge in a conduit are then related through (Fig. 1d)

$$N^n = \frac{c_1 Q \Psi + u_b h}{c_2 c_3^{-1/\alpha} Q^{1/\alpha} \Psi^{-1/(2\alpha)}} \tag{2}$$

At low discharge, Q, the effective pressure, N, decreases with Q, as is expected for cavities, whereas at higher discharge, N increases with Q and the conduit behaves as a Röthlisberger channel. The switch-over in behaviour occurs at a critical discharge

$$Q_c = \frac{u_b h}{c_1 (\alpha - 1) \Psi}$$

Below Q_c, the conduit is kept open mainly by ice flow over bed protrusions; above Q_c, it is kept open by wall melting.

A linear stability analysis (Supplementary Information) also shows that discharge becomes concentrated into a few conduits when the mean water discharge through an array of laterally connected conduits exceeds Q_c: driven by wall melting, a single conduit will grow into a large channel (with the properties of a Röthlisberger channel, its size, S,

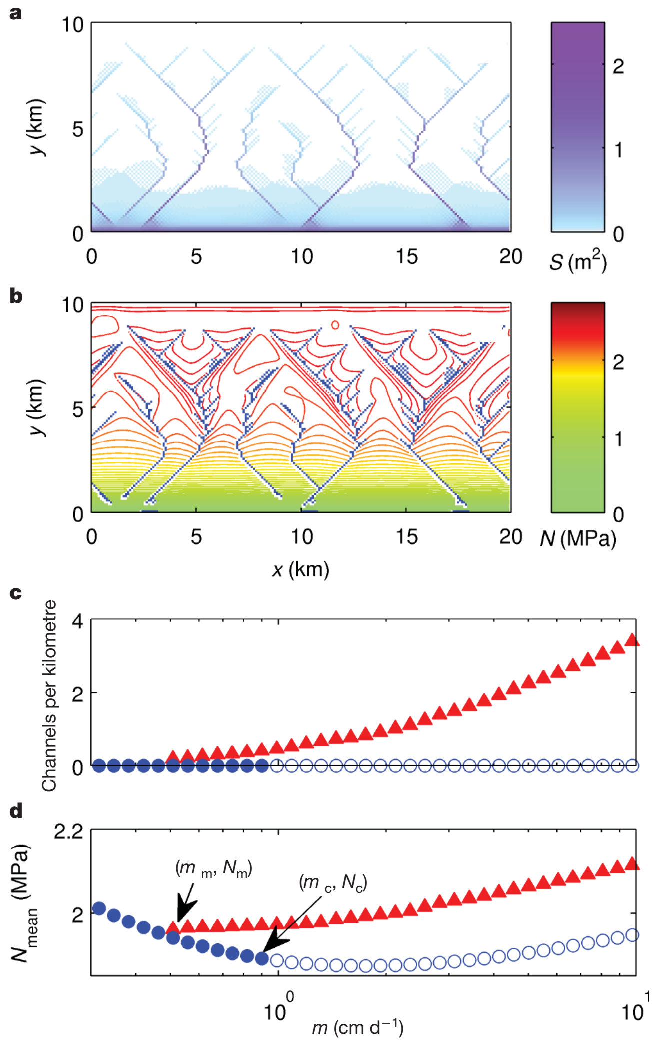


Figure 2 | Steady-state drainage systems. **a**, **b**, Example of a drainage system formed spontaneously through the channelizing instability. **a**, Conduit sizes. Channels are much larger (dark blue and purple) than the surrounding cavities. **b**, Channels are shown in blue and effective pressure contours are shown at 0.05-MPa intervals. The pressure distribution reveals how channel–cavity interactions control the drainage pattern. Channels are at higher effective pressure than the surrounding cavities. Local water pressure maxima (minima of N) separate the channels, driving water flow towards them. **c**, **d**, Steady-state drainage system characteristics as functions of water supply rate, m. **c**, Channel density (average number of channels per unit width of the domain) plotted against m. **d**, Mean of N over the domain plotted against m. Red triangles correspond to channelized systems; blue circles correspond to unchannelized ones. Open circles show unstable unchannelized systems (which will evolve into a channelized state if perturbed). Instability first occurs at a critical water supply, m_c, corresponding to a critical discharge, Q_c. Mean effective pressure decreases with water supply (and, hence, discharge) for stable unchannelized systems, and increases with water supply for channelized ones. For some intermediate values of m (between m_c and a lower limit, m_m, that corresponds to a critical lower discharge, Q_m), both channelized and unchannelized states are possible: their low water pressure allows channels to suck in enough water to keep themselves open, but the discharge through the system is too low for an unchannelized system to channelize spontaneously. A video animation is included in Supplementary Information.

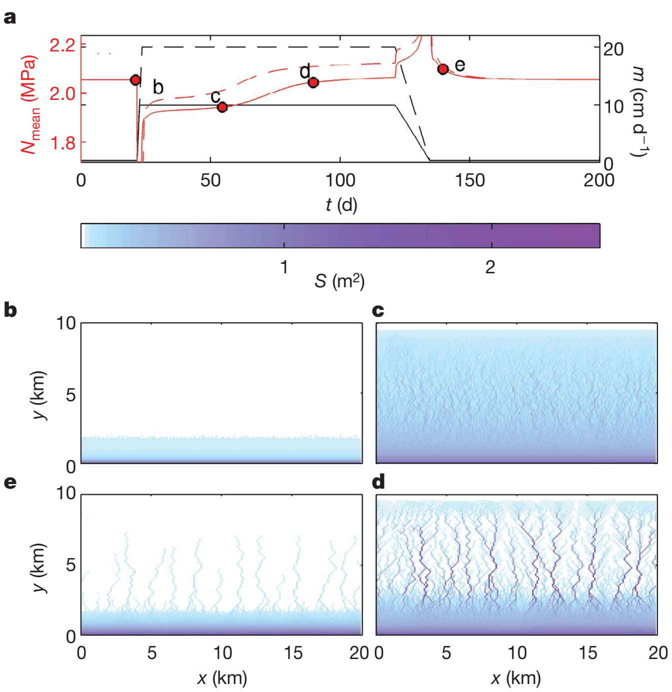


Figure 3 | Idealized seasonal evolution of the drainage system. **a**, The spatial mean of effective pressure, N (red lines), plotted against time. The simulations shown are forced by a sharp increase (over 1 d) in water supply, m (black line), from a wintertime value of 0.33 cm d^{-1} to a summertime value of 10 cm d^{-1} (solid lines) and 20 cm d^{-1} (dashed lines). This is followed by steady supply for 100 d and a gradual return to 0.33 cm d^{-1}. The dots marked b–e correspond to the spatial drainage configurations shown in panels **b**–**e**, respectively. **b**–**e**, The drainage system starts close to an unchannelized steady state with small conduits (**b**). The abrupt increase in m leads to a sharp drop in effective pressure (a 'spring event'[9]), which opens the drainage conduits to accommodate the additional discharge but does not immediately channelize the system (**c**). Efficient channelization causes effective pressure to increase only after some time (**d**), reaching values above those of wintertime. The final drop in m causes a temporary jump in effective pressure that leads the system to shut down for winter (**e**). Both simulations in panel **a** show qualitatively the same response. However, the larger jump in water supply (dashed lines in **a**) leads to a shorter and less pronounced period of low effective pressure than the smaller jump (solid lines in **a**). A video animation is included in Supplementary Information.

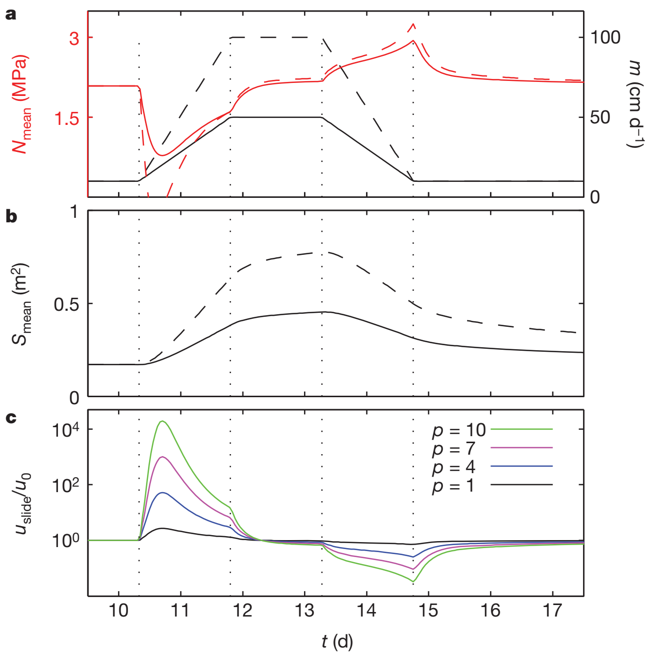


Figure 4 | Temporal variations in water input. **a**, The mean of N (red) over the domain plotted against time for two different simulations. Simulations are started from a steady-state channelized system and forced with time-dependent but spatially uniform water input, m (black), imposing fivefold (solid lines) and tenfold (dashed lines) increases in m over 4 d. **b**, The spatial mean of conduit size, S, plotted against time. During the initial increase in water input, conduits have not yet been able to widen to accommodate increased discharge. To force the additional discharge instead requires a temporary spike in hydraulic gradient, Ψ, leading to higher water pressure (lower N) upstream of the margin (red lines). This temporary drop in N is stronger for bigger jumps in m (dashed lines in **a**): N_{mean} can even drop to zero, which corresponds to complete decoupling between ice and bed. Hydrofracture should occur[30], although this is not included in my model. After the initial transient, conduit size adjusts and effective pressure increases again, reaching a maximum when m decreases again. **c**, Modelled sliding velocity, u_{slide}, normalized by steady-state sliding velocity, u_0. Time series of u_{slide}/u_0 are shown corresponding to the solid curves in **a** and **b**. Sliding is modelled using the empirical relation[9,23] $\tau_b = Cu_{slide}^{1/p}N$, where τ_b is driving stress in the ice and C and p are constant parameters (Supplementary Information). The curves correspond to different values of the sliding-law nonlinearity, p, as indicated. In all cases, the initial drop in N leads to fast sliding. Recent developments[10,11] in glacier sliding suggest large values for p, for which the magnitude of sliding events is more pronounced. The calculation for u_{slide}, however, excludes the effects of stress transfer to other parts of the glacier, which would prevent excessively large sliding velocities.

and effective pressure, N, increasing with discharge, Q) at the expense of nearby ones, which shrink to form smaller cavities. Below this critical mean discharge, all conduits can be stable at the same size and behave as cavities (in which the steady-state effective pressure decreases with increasing discharge).

The nonlinear dynamics of channelization can be captured by considering a network[26] of conduits described by equation (1) (Methods and Supplementary Information). With mean discharge below a critical value, Q_c, an initially nearly uniform network remains uniform as predicted by linear stability analysis. For a mean discharge level greater than Q_c, the channelizing instability occurs and the system spontaneously evolves a set of large, well-defined channels fed by smaller ones that are separated in turn by cavities (Fig. 2). This effect is similar to melt channelization in magmatic systems[27]. The spacing between the channels is controlled by lateral effective pressure gradients and decreases with increasing water input. An important feature of the nonlinear system is that channelization is irreversible. Even if the mean discharge is dropped back below Q_c, the previously formed channels do not necessarily disappear: this requires discharge to drop below a lower critical level, Q_m (Fig. 2).

An increase in steady meltwater supply lowers the effective pressure and therefore speeds up sliding[9–11] only below the critical discharge, Q_c, for channelization (equation (2); Figs 1d and 2d). Once this is exceeded, the effective pressure increases again. Channelization increases the effective pressure further: concentrated discharge leads to faster channel wall melting that must be offset by stronger creep closure, driven by increased N. An increase in steady meltwater input therefore has limited potential to cause glacier acceleration and will eventually even lead to glacier deceleration.

This result, however, applies only to steady conditions. Observations in Greenland[6] indicate that seasonal and short-term water supply variations can lead to transient acceleration. Ice velocities in some areas are consistently above their wintertime average early in the melt season, but slow down to below their wintertime average later in summer. This can be explained by a seasonal switch from unchannelized to channelized drainage, in which a combination of increased water supply and incomplete channelization cause low effective pressures in early summer (Fig. 3). However, Fig. 3 also shows that channelization occurs faster and that the decrease in effective pressures early in the melt season is smaller when summertime water supply rates are large. Higher summer surface melt rates are therefore likely to suppress the magnitude and duration of the period of higher velocities in early summer.

Short-term spikes in water supply can also induce spikes in water pressure, and lead to the observed[6] short-term ($\lesssim$1-day) fast-sliding episodes even when the drainage system has channelized[13] (Fig. 4). This happens because the size of conduits adjusts slowly (over several days), and the drainage system does not have the capacity to accommodate sudden extra water throughput except by an increase in the hydraulic gradient, Ψ. This increase in Ψ requires higher water pressures in the interior of the drainage system, leading to lower effective pressures and, hence, to faster sliding. Not only can short-term variability lead to acceleration even after channelization, but the magnitude of water pressure excursions during short-term water supply spikes can also be much larger than the slower seasonal water pressure signal (compare Figs 3a and 4a).

Ice velocity can therefore respond much more to short-term temporal variations in water supply than to changes in mean water flow. This has major implications for ice-sheet dynamics and feedbacks between surface melting and dynamic thinning[7]. More surface water input through melt or rain is likely if dynamic thinning draws down the ice surface. This can lead to increased ice flow and further thinning if the basal water supply is initially very low or if the bed is frozen. However, larger rates of summer water supply can also cause faster channelization and potential ice deceleration. Further acceleration must then be driven instead by short-term temporal variability in water supply. This is favoured by strong diurnal cycles[5] and frequent rain events[17], both of which are more likely at lower latitudes, or if the ice sheet develops numerous surface lakes that drain abruptly[18].

Drainage channelization under glaciers and ice sheets suppresses the ability of steady surface water supply to cause further ice acceleration, but faster ice flow can be caused instead by water input variations[8]. This is already observable in Greenland and will become more important when the climate changes: diurnal melt cycles already contribute to ice flow in southern Greenland[5], and more frequent rain events are predicted to result from a northward shift of storm tracks over the next century[28], which will cause further ice acceleration. My results are also relevant to palaeo-ice-sheet dynamics. Simulations that do not include subglacial processes cannot explain the observed rapid collapse of the Laurentide ice sheet[29]. A water input/dynamic thinning feedback is a plausible collapse mechanism, driven by rain and diurnal melt cycles rather than by mean melt alone. Future coupled models are needed to fully explain the role of drainage in rapid deglaciation, and my results show that channelization and short-term drainage variability are the crucial processes that must be captured in these models.

METHODS SUMMARY

I model a drainage network in which nodes i and j are connected by a conduit (network edge) labelled by subscripts i and j. The conduit evolves according to

$$\frac{dS_{ij}}{dt} = c_1 Q_{ij} \Psi_{ij} + u_b h - c_2 N_{ij}^n S_{ij}$$

where S_{ij}, Q_{ij}, Ψ_{ij} and N_{ij} are conduit size, flux, hydraulic gradient and effective pressure, respectively. Conduit sizes, S_{ij}, and effective pressures at the nodes, N_i, are the primary variables. I set $\Psi_{ij} = \Psi_{ij}^0 + (N_j - N_i)/L_{ij}$, where L_{ij} is the distance between nodes and Ψ_{ij}^0 is a geometrically controlled background hydraulic gradient (Supplementary Information). Additionally, $N_{ij} = (N_i + N_j)/2$ and $Q_{ij} = c_3 S_{ij}^{\alpha} |\Psi_{ij}|^{-1/2} \Psi_{ij}$. At each node, mass is conserved. Ignoring water storage (Supplementary Information), mass conservation requires that

$$\sum_j Q_{ij} = m_i$$

where the sum is over the nodes, j, connected to the node i, and m_i is water input to node i. I use a rectangular lattice network oriented at 45° to downslope, with a domain size of 10 km × 20 km and 2×10^4 conduits. I impose $N = 0$ at the margin, zero inflow upstream and periodic sides. Water input is spatially uniform (all m_i are the same) and is given as rate of volume input per unit area. The parameters are $\alpha = 5/4$, $c_1 = 3.4 \times 10^{-9}\ \mathrm{Pa}^{-1}$, $c_2 = 4.5 \times 10^{-25}\ \mathrm{Pa}^{-3}\ \mathrm{s}^{-1}$, $c_3 = 0.33\ \mathrm{kg}^{-1/2}\ \mathrm{m}^{3/2}$ and $u_b h = 3\ \mathrm{m}^2\ \mathrm{yr}^{-1}$ (Supplementary Information). For illustrative purposes, Ψ_{ij}^0 is based on the shape of a plastic glacier with a yield stress of 10^5 Pa on a 3° slope. In Fig. 1 $\Psi = 512\ \mathrm{Pa\ m}^{-1}$ and $u_b h = 3\ \mathrm{m}^2\ \mathrm{yr}^{-1}$, and in Fig. 1c $N = 2.85$ MPa.

Received 26 April; accepted 25 October 2010.

1. Rignot, E. & Kanagaratnam, P. Changes in the velocity structure of the Greenland ice sheet. *Science* **311,** 986–990 (2006).
2. Joughin, I. *et al.* Seasonal speedup along the western flank of the Greenland ice sheet. *Science* **320,** 781–783 (2008).
3. Zwally, H. J. *et al.* Surface-melt induced acceleration of Greenland ice-sheet flow. *Science* **5579,** 218–222 (2002).
4. van de Wal, R. S. W. *et al.* Large and rapid melt-induced velocity changes in the ablation zone of the Greenland ice sheet. *Science* **321,** 111–113 (2008).
5. Shepherd, A. *et al.* Greenland ice sheet motion coupled with daily melting in late summer. *Geophys. Res. Lett.* **36,** L01501 (2009).
6. Bartholomew, I. *et al.* Seasonal evolution of subglacial drainage and acceleration in a Greenland outlet glacier. *Nature Geosci.* **3,** 408–411 (2010).
7. Parizek, B. R. & Alley, R. B. Implications of increased Greenland surface melt under global-warming scenarios: ice-sheet simulations. *Quat. Sci. Rev.* **23,** 1013–1027 (2004).
8. Bartholomaus, T. C., Anderson, R. S. & Anderson, S. P. Response of glacier basal motion to transient water storage. *Nature Geosci.* **1,** 33–37 (2008).
9. Iken, A. & Bindschadler, R. A. Combined measurements of subglacial water pressure and surface velocity of Findelengletscher, Switzerland: conclusions about drainage system and sliding mechanism. *J. Glaciol.* **32,** 101–119 (1986).
10. Iverson, N. R., Baker, R. W., Hooke, R. LeB., Hanson, B. & Jansson, P. Coupling between a glacier and a soft bed: I. A relation between effective pressure and local shear stress determined from till elasticity. *J. Glaciol.* **45,** 31–40 (1999).
11. Schoof, C. The effect of cavitation on glacier sliding. *Proc. R. Soc. Lond. A* **461,** 609–627 (2005).
12. Kamb, B. *et al.* Glacier surge mechanism: 1982–1983 surge of Variegated Glacier, Alaska. *Science* **227,** 469–479 (1985).
13. Iken, A., Echelmeyer, K. A., Harrison, W. D. & Funk, M. Mechanisms of fast flow in Jakobshavns Isbrae, Greenland, part I: measurements of temperature and water level in deep boreholes. *J. Glaciol.* **39,** 15–25 (1993).
14. Hubbard, B., Sharp, M. J., Willis, I. C., Nielsen, M. K. & Smart, C. C. Borehole water-level variations and the structure of the subglacial hydrological system of Haut Glacier d'Arolla, Valais, Switzerland. *J. Glaciol.* **41,** 572–583 (1995).
15. Lappegard, G., Kohler, J., Jackson, M. & Hagen, J. O. Characteristics of subglacial drainage system deduced from load-cell measurements. *J. Glaciol.* **52,** 137–147 (2006).
16. Magnusson, E., Björnsson, H., Rott, H. & Palsson, F. Reduced glacier sliding caused by persistent drainage from a subglacial lake. *Cryosphere* **4,** 13–20 (2010).
17. Howat, I. M., Tulaczyk, S., Waddington, E. & Björnsson, H. Dynamic controls on glacier basal motion inferred from surface ice motion. *J. Geophys. Res.* **113,** F03015 (2008).
18. Das, S. B. *et al.* Fracture propagation to the base of the Greenland ice sheet during supraglacial lake drainage. *Science* **320,** 778–781 (2008).
19. Röthlisberger, H. Water pressure in intra- and subglacial channels. *J. Glaciol.* **11,** 177–203 (1972).
20. Nye, J. F. Water flow in glaciers: jökulhlaups, tunnels and veins. *J. Glaciol.* **17,** 181–207 (1976).
21. Walder, J. Hydraulics of subglacial cavities. *J. Glaciol.* **32,** 439–445 (1986).
22. Kamb, B. Glacier surge mechanism based on linked cavity configuration of the basal water conduit system. *J. Geophys. Res.* **92,** 9083–9100 (1987).
23. Fowler, A. C. Sliding with cavity formation. *J. Glaciol.* **33,** 255–267 (1987).
24. Lüthi, M., Funk, M., Iken, A., Gogineni, S. & Truffer, M. Mechanisms of fast flow in Jakobshavns Isbrae, Greenland, part III: measurements of ice deformation, temperature and cross-borehole conductivity in boreholes to the bedrock. *J. Glaciol.* **48,** 369–385 (2002).
25. Hewitt, I. J. & Fowler, A. C. Seasonal waves on glaciers. *Hydrol. Process.* **22,** 3919–3930 (2008).
26. Clarke, G. K. C. Lumped-element analysis of subglacial hydraulic circuits. *J. Geophys. Res.* **101,** 17547–17559 (1996).
27. Hewitt, I. J. & Fowler, A. C. Melt channelization in ascending mantle. *J. Geophys. Res.* **114,** B06210 (2009).
28. Schuenemann, K. C. & Cassano, J. J. Changes in synoptic weather patterns and Greenland precipitation in the 20th and 21st centuries: 2. Analysis of 21st century atmospheric changes using self-organizing maps. *J. Geophys. Res.* **115,** D05108 (2010).
29. Tarasov, L. & Peltier, W. R. Terminating the 100 kyr ice age cycle. *J. Geophys. Res.* **102,** 21665–21693 (1997).
30. Tsai, V. C. & Rice, J. R. A model for turbulent hydraulic fracture and application to crack propagation at glacier beds. *J. Geophys. Res.* **115,** F03007 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Thanks to G. Clarke, T. Creyts, G. Flowers, I. Hewitt, R. Hindmarsh, M. Jellinek and V. Radić for comments on the manuscript, and to C. Koziol and M. Jaffrey for discussions. Financial support was provided by the Canada Research Chairs Program, NSERC Discovery Grant 357193-08 and the Canadian Foundation for Climate and Atmospheric Science through the Polar Climate Stability Network.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to the author (cschoof@eos.ubc.ca).

doi:10.1038/nature09636

Outer-core compositional stratification from observed core wave speed profiles

George Helffrich[1]† & Satoshi Kaneshima[2]

Light elements must be present in the nearly pure iron core of the Earth to match the remotely observed properties of the outer and inner cores[1,2]. Crystallization of the inner core excludes light elements from the solid, concentrating them in liquid near the inner-core boundary that potentially rises and collects at the top of the core[3], and this may have a seismically observable signal. Here we present array-based observations of seismic waves sensitive to this part of the core whose wave speeds require there to be radial compositional variation in the topmost 300 km of the outer core. The velocity profile significantly departs from that of compression of a homogeneous liquid. Total light-element enrichment is up to five weight per cent at the top of the core if modelled in the Fe–O–S system. The stratification suggests the existence of a subadiabatic temperature gradient at the top of the outer core.

Many light elements, namely hydrogen, carbon, nitrogen, oxygen, sulphur and silicon, could plausibly have been included in the core during accretion of the early Earth[2,4]. Among these, oxygen and sulphur are of interest owing to oxygen's potentially high solubility in iron at high pressures and the extremely low-temperature eutectic between iron and sulphur, which facilitates segregation into the core[5–8]. Independently of the particular light elements involved, their enrichment in the liquid of the outer core is potentially observable because it affects liquid density and bulk modulus, thereby changing seismic wave speeds. Indeed, a series of past seismic studies focused on this area specifically to seek evidence for stratification[9–12]; many suggest lower seismic wave speeds in the outer core near the core–mantle boundary (CMB) than self-compression of a chemically homogeneous outer core would imply.

Our study uses earthquakes in South America and in the southwestern Pacific region that emit shear waves towards the core and convert to compressional waves that repeatedly reflect from the underside of the CMB (Fig. 1). The arrivals, S*m*KS, reflect $m-1$ times from the core side of the CMB. Thus most of their travel time is accrued across the area of potential light-element accumulation, between 80 and 400 km below the CMB. We use stacked records of 120–190 individual observations of S*m*KS waveforms of three separate earthquakes recorded by large-scale seismic arrays in Japan and northern Europe to measure differential travel times and slownesses between S*m*KS and S*n*KS, with $m > n$, $2 \leq (m \text{ and } n) \leq 5$ (Fig. 1). The events are selected for their range (separating S*m*KS multiples with $m \geq 4$) and for their sampling of different surface and CMB environments. Our array-based measurements on single events yield stacked waveforms whose differential times and slownesses are the raw data used. Array measurements provide not only direct differential slowness information; more importantly, the stacked waveforms average over source- and receiver-side near-CMB path effects due to D″ structure.

In the model, named KHOCQ, derived from the observations, we find that outer-core wave speeds decrease to values a maximum of 0.3% lower than PREM wave speeds[13] 60 km into the outer core and gradually return to PREM values at a depth of ~300 km (Fig. 2). These small, resolvable differences are significant departures from homogeneous self-compression of core material[1]. For two regions with very different seismic velocity structures in the lowermost mantle, we find a thicker anomalous layer (300 km) than any of the earlier studies, owing to the use of *S4KS* and *S5KS*, which is particularly suitable for investigating the outermost core but has not yet been extensively used. Isolation of propagation delays accrued at shallowest core levels improves resolution of deeper structure.

To interpret the observations, we model seismic wave speeds in Fe–O–S liquids under core conditions[14] (Methods). Given composition, pressure and temperature, the model provides liquid bulk modulus, K, density, ρ, and, thus, seismic wave speed, $(K/\rho)^{1/2}$. We calculate compositional profiles at the top of the outer core that match our wave speed profile to within 0.02%, have average densities in the top 300 km of the outer core within 1% of PREM (and thereby obey normal-mode-derived density constraints[15]), and that merge with PREM wave speeds 300 km below the CMB. Feasible variations in iron, oxygen and sulphur that lead to stable stratification are shown in Fig. 3. The modelled compositions agree with KHOCQ for oxygen enrichments towards the CMB of 0.8–2 wt% and iron depletions of up to 5 wt%; sulphur may be either enriched by up to 3 wt% or depleted by 0.8 wt%, depending on the oxygen content of the liquid. At light-element enrichment of 5% at the top of the core, the total added light-element mass in the region is about 11% of the inner core's mass. We can equate the light-element mass expelled by inner-core crystallization with light-element enrichment in the layer by assuming that the density jump at the inner-core boundary, $\Delta\rho_{\mathrm{ICB}}$, multiplied by the inner core's present volume gives this mass[14]. Using a previous estimate[15] of $610 \pm 180\ \mathrm{kg}\ m^{-3}$ for $\Delta\rho_{\mathrm{ICB}}$ attributable to light-element release, we find the light-element mass in the layer agrees (at the 95% confidence level) with the mass expelled by the inner core if enrichment is 2.5–3 wt% at the top of the outer core. The range of seismically defined $\Delta\rho_{\mathrm{ICB}}$ estimates is quite wide, however, so the rough mass balance demonstrates the feasibility of the layer's formation mechanism rather than providing strong compositional constraints.

The lighter material at the top of the core is 5.9% less dense than the liquid 300 km below it (and 1.6% lighter than PREM densities on the core side of the CMB). The layer's Brunt–Väisälä frequency, a measure of the strength of density stratification in the core[16], is 0.51–1.03 mHz (periods of 1.63–3.43 h), suggesting strong stabilization. The density excess in the layer, -0.8×10^{-2} to -3.2×10^{-2}, is about 100 times greater than that estimated[17] to explain short-period geomagnetic field secular variation and its possible correlation with length-of-day variations. The layer's density gradient exceeds that in the well-mixed outer core[16] by a factor exceeding 10^{10}. Owing to the gradational difference in composition from the bulk of the outer core at the layer's base, stabilizing it against convective mixing in the outer core is problematic (see Supplementary Information for a discussion). A feasible mechanism is a subadiabatic gradient at the top of the outer core due to the inability of core heat to escape into the mantle on account of its slower convection speeds and low thermal diffusivity[18,19]. The core liquid model (Methods) yields velocities that match our observed profiles

[1]Earthquake Research Institute, University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan. [2]Earth and Planetary Sciences, University of Kyushu, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan. †Present address: Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol BS8 1RJ, UK.

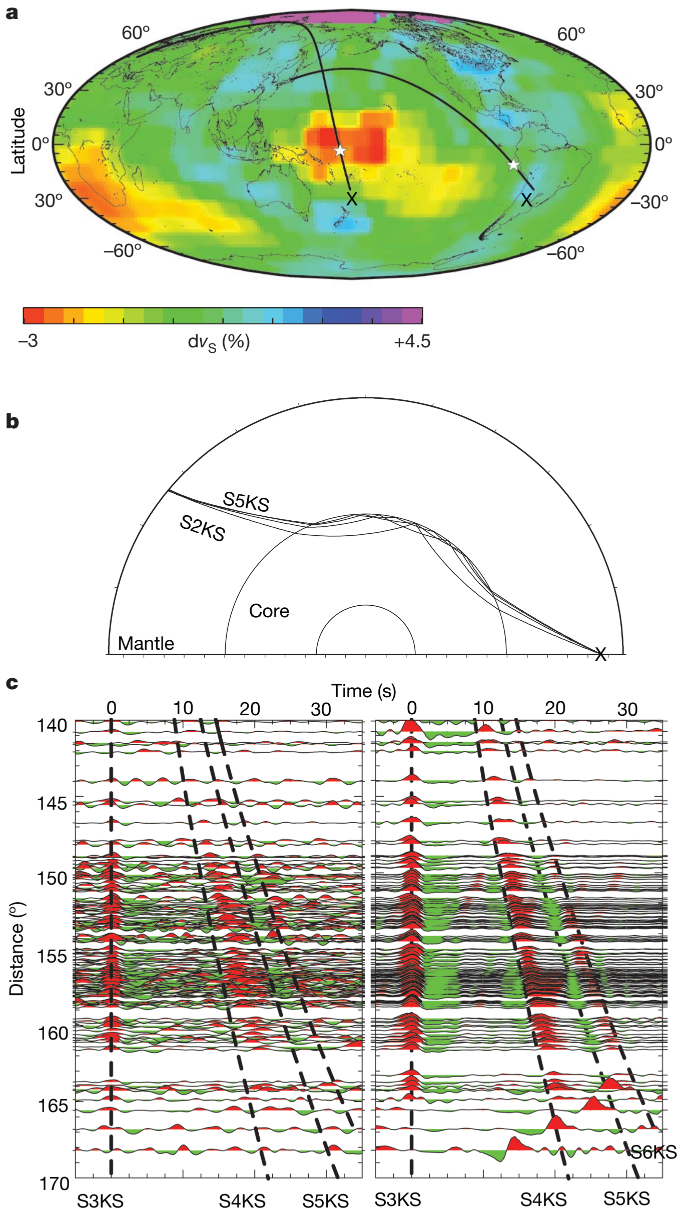


Figure 1 | Experiment geometry and paths taken by the seismic waves through the Earth. **a**, Sources (X) are subduction-zone earthquakes in South America (Argentina) and the southwest Pacific Ocean (Fiji) recorded by seismic arrays in Japan (~120 stations) and in Europe (~190 stations). The ray paths are superimposed on tomography maps (SB10L18) of shear-wave speed variations, dv_S, at 2,770 km in the lowermost mantle[26]. White stars indicate representative core entry points of S3KS. **b**, S*m*KS ray paths travel across the mantle as shear waves, but across the core as compressional waves. In the core, they reflect $m-1$ times from the underside of the CMB. **c**, Record section of observed (left, recorded by European stations) and predicted (right) S*m*KS arrivals from the Fiji earthquake. Arrivals and synthetics are aligned on S3KS (0 s), and the Preliminary Reference Earth Model[13] (PREM) predicted successive S*m*KS arrivals shown by dashed lines. Reflectivity synthetics for PREM (right) show that S4KS and S5KS are delayed. S2KS, which arrives before S3KS, is not shown, for clarity. PcPPKiKS is the prominent low-slowness arrival in the synthetics at 163–170°, not observable in the data.

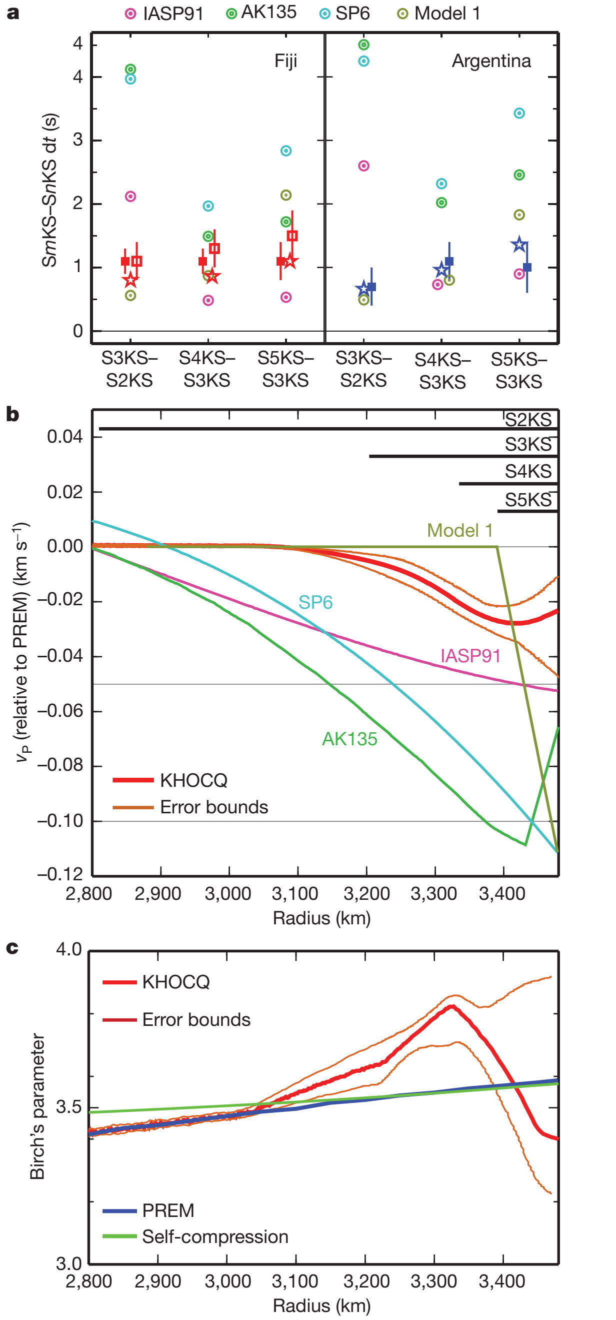


Figure 2 | Data, velocity profile in the outer core and self-compression profile using the velocity profile. **a**, Raw data are differential travel times (d*t*) and slownesses measured relative to PREM[13] for two regions. Data (and 95% observational uncertainty) for the two Fiji events and one Argentina event studied are shown with squares, previous model predictions (see **b**) are shown with circles and KHOCQ predictions are shown with stars. Most models significantly overestimate differential travel times relative to observations. **b**, Outer-core velocity resulting from τ–p inversion of differential S*m*KS travel times and slownesses (Methods). Velocities are shown relative to PREM[13] and are compared with recent models incorporating similar S*m*KS data: IASP91[27], SP6[28], AK135[29] and model 1 from ref. 11. Lines across the top indicate depths into the core travelled by each S*m*KS path. Higher multiples travel to smaller depths in the outer core. The KHOCQ error bounds correspond to the range of inversions with 2σ travel-time uncertainties applied—roughly the 95% confidence level. **c**, Birch's parameter (and uncertainty bounds) calculated from KHOCQ departs significantly from self-compression in the range 80–300 km below the CMB, requiring compositional change in the outer-core liquid. The self-compression line shows the theoretical behaviour of a substance whose composition does not vary (Methods). Birch's parameter for PREM closely approximates that for self-compression.

either for adiabatic temperature gradients in the topmost 300 km of the core or for an isothermal layer for temperatures between 3,000 and 5,500 K, so the results are independent of the temperature structure chosen. Higher temperatures imply a bulk core composition more iron rich by 1–2 wt% and correspondingly larger light-element enrichments. This trade-off between core temperature and light-element

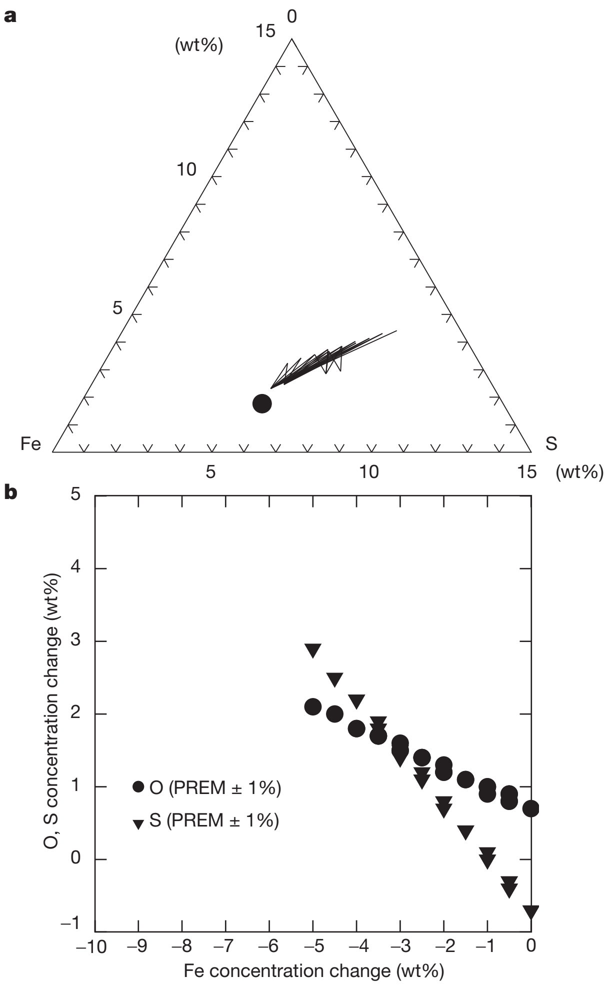


Figure 3 | Compositional variation in the Fe–O–S liquids that match the observed wave speed profile and core density constraints. **a**, Feasible composition profiles are shown by lines linking PREM wave speeds at 300-km depth in the outer core to wave speeds at the top of the outer core. Profiles are isothermal in the topmost 300 km and then join an adiabat initiated at the CMB at 4,300 K. For feasibility, core liquid compositions must be within 1% of the average PREM density in the top 300 km of the core[15] and must match the observed velocity profile. For comparison with previous core composition estimates, the filled circle indicates a first-principles core liquid composition estimate[30]. **b**, Compositional variation along feasible profiles for 1% density agreement with PREM. The compositional difference at the CMB and the base of each feasible profile shown in **a** (presumably the well-mixed outer core represented by PREM wave speeds 300 km below the CMB) is shown here. Oxygen enrichment by >0.8 wt% is required in all cases and sulphur enrichment is required only at oxygen contents ≥1 wt%.

concentration also adds uncertainty to the light-element balance between the inner core and the layer discussed earlier.

The liquid model we use only approximates the true composition of the core because it includes only iron and accounts for only oxygen and sulphur as light elements. Nickel is also thought to be present in the core, in concentrations up to ~5 wt% (ref. 4). However, in these concentrations it seems to act equivalently to iron in sulphur- and oxygen-bearing metallic liquids, in the sense that melt surface energies are unchanged when nickel replaces an equivalent amount of iron[20]. Furthermore, (Fe,Ni)–(Fe,Ni)S eutectic temperatures are only 20 °C lower than Fe–FeS (ref. 21). Thus, except for the minor difference in density (<0.1%) due to the substitution of nickel for iron, we expect no significant effects on liquid properties. Present cosmochemical models for core composition suggest that silicon might also be present in concentrations of a few weight per cent[4]. At present, we lack the requisite melting data on Fe–Si alloys to develop an analogous model for silicon-bearing melts. We suspect that sulphur might decrease in proportion to silicon's increase[22] and lead to a slight decrease in density (<0.1%, due to the slight atomic weight difference), but the bulk modulus effect on wave speeds cannot yet be assessed owing to the lack of Fe–Si melting data.

If stabilized subadiabatically against convection, core–mantle reaction might inject oxygen into the outer core[5,7,23]. Oxygen diffusion from the CMB cannot be responsible for the complete depth range of the observed anomaly, however. Anion diffusivities in liquid iron are estimated to be 10^{-11}–10^{-9} $m^2 s^{-1}$ (refs 3, 24), leading to diffusion length scales of 10–100 km over the lifetime of the Earth. Thus, the profile was in large part created by an upward buoyant flux of light elements from the crystallization of the inner core. The outer core's velocity profile seems to record the secular evolution of the core's composition as the subadiabatic layer grew, possibly modified by core–mantle reaction. We remark in passing that a subadiabatic layer at the top of the core also has profound implications for the thermal evolution of the core and CMB thermal structure.

METHODS SUMMARY

Data are broadband records from regional networks in Europe and Japan. The radial component waveforms are uncorrected for instrument response and correspond to ground velocity given typical instrument characteristics (Supplementary Information). We measured differential travel times and slownesses of S3KS–S2KS, S4KS–S3KS and S5KS–S3KS on stacked and Hilbert-transformed waveforms (when required). These are the raw data for the inversion.

We determined velocity profiles by numerical τ–p inversion[25] using up to six basis functions, yielding flattened earth velocity perturbations, $\delta v_f/v_f$, of the form $(\delta v_f/v_f)(r) = \pm 0.01\max^2(0, r - r_b)/(r_{CMB} - r_b)^2$, where r is radius and r_b and r_{CMB} are the minimum perturbation radius and the CMB radius, respectively. See Supplementary Information for fitting details and coefficients. The Fe–O–S liquid model is from ref. 14 and is based on thermodynamic fits to the melting curves of Fe, FeO and FeS using a 1-bar metallurgical model of liquid free energy. See Methods for the fitting procedure and computational methods; liquid thermophysical properties are listed in Supplementary Information. The model yields high-pressure and -temperature bulk modulus and density, and, therefore, seismic wave speeds in outer-core liquid. Because the model is free-energy based, the liquid heat capacity, thermal expansivity and, thus, the adiabatic gradient in the liquid can also be calculated through thermodynamic identities relating to free-energy derivatives.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 May; accepted 29 October 2010.

1. Birch, F. Elasticity and constitution of the Earth's interior. *J. Geophys. Res.* **57,** 227–286 (1952).
2. Stevenson, D. J. Models of the Earth's core. *Science* **214,** 611–619 (1981).
3. Fearn, D. R. & Loper, D. E. Compositional convection and stratification of Earth's core. *Nature* **289,** 393–394 (1981).
4. Wood, B. J., Walter, M. J. & Wade, J. Accretion of the Earth and segregation of its core. *Nature* **441,** 825–833 (2006).
5. Rubie, D. C., Gessmann, C. K. & Frost, D. J. Partitioning of oxygen during core formation on the Earth and Mars. *Nature* **429,** 58–61 (2004).
6. Ozawa, H. *et al.* Chemical equilibrium between ferropericlase and molten iron to 134 GPa and implications for iron content at the bottom of the mantle. *Geophys. Res. Lett.* **35,** L05308 (2008).
7. Frost, D. J. *et al.* Partitioning of oxygen between the Earth's mantle and core. *J. Geophys. Res.* **115,** B02202 (2010).
8. Hilty, D. C. & Crafts, W. Liquidus surface of the Fe-S-O system. *J. Metals (Trans. AIME)* **4,** 1307–1312 (1952).
9. Lay, T. & Young, C. The stably-stratified outermost core revisited. *Geophys. Res. Lett.* **17,** 2001–2004 (1990).
10. Garnero, E. J., Helmberger, D. V. & Grand, S. P. Constraining outermost core velocity with SmKS waves. *Geophys. Res. Lett.* **20,** 2463–2466 (1993).
11. Tanaka, S. Possibility of a low P-wave velocity layer in the outermost core from global SmKS waveforms. *Earth Planet. Sci. Lett.* **259,** 486–499 (2007).
12. Alexandrakis, C. & Eaton, D. W. Precise seismic-wave velocity atop Earth's core: no evidence for outer-core stratification. *Phys. Earth Planet. Inter.* **180,** 59–65 (2010).
13. Dziewonski, A. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter.* **25,** 297–356 (1981).
14. Helffrich, G. & Kaneshima, S. Seismological constraints on core composition from Fe-O-S liquid immiscibility. *Science* **306,** 2239–2242 (2004).
15. Masters, G. & Gubbins, D. On the resolution of density within the Earth. *Phys. Earth Planet. Inter.* **140,** 159–167 (2003).
16. Loper, D. E. A model of the dynamical structure of the Earth's outer core. *Phys. Earth Planet. Inter.* **117,** 179–196 (2000).

17. Braginsky, S. I. MAC-oscillations of the hidden ocean of the core. *J. Geomag. Geoelectr.* **45,** 1517–1538 (1993).
18. Gubbins, D., Thomson, C. & Whaler, K. Stable regions in the Earth's liquid core. *Geophys. J. R. Astron. Soc.* **68,** 241–251 (1982).
19. Lister, J. R. & Buffett, B. A. Stratification of the outer core at the core-mantle boundary. *Phys. Earth Planet. Inter.* **105,** 5–19 (1998).
20. Rose, L. A. & Brenan, J. M. Wetting properties of Fe-Ni-Co-Cu-O-S melts against olivine: implications for sulfide melt mobility. *Econ. Geol.* **96,** 145–157 (2001).
21. Usselman, T. M. Experimental approach to the state of the core: part I. The liquidus relations of the Fe-rich portion of the Fe-Ni-S system from 30 to 100 kb. *Am. J. Sci.* **275,** 278–290 (1975).
22. Kilburn, M. R. & Wood, B. J. Metal-silicate partitioning and the incompatibility of S and Si during core formation. *Earth Planet. Sci. Lett.* **152,** 139–148 (1997).
23. Buffett, B. A. & Seagle, C. T. Stratification of the top of the core due to chemical interactions with the mantle. *J. Geophys. Res.* **115,** B04407 (2010).
24. Dobson, D. Self-diffusion in liquid Fe at high pressure. *Phys. Earth Planet. Inter.* **130,** 271–284 (2002).
25. Garmany, J., Orcutt, J. A. & Parker, R. L. Travel time inversion: a geometrical approach. *J. Geophys. Res.* **84,** 3615–3622 (1979).
26. Masters, G., Laske, G., Bolton, H. & Dziewonski, A. in *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale* (eds Karato, S.-I., Forte, A. M., Liebermann, R. C., Masters, G. & Stixrude, L.) 63–87 (Geophys. Monogr. 117, American Geophysical Union, 2000).
27. Kennett, B. L. N. & Engdahl, E. R. Traveltimes for global earthquake location and phase identification. *Geophys. J. Int.* **105,** 429–465 (1991).
28. Morelli, A. & Dziewonski, A. M. Body wave traveltimes and a spherically symmetric P- and S-wave velocity model. *Geophys. J. Int.* **112,** 178–194 (1993).
29. Kennett, B. L. N., Engdahl, E. R. & Buland, R. Constraints on seismic velocities in the Earth from traveltimes. *Geophys. J. Int.* **126,** 108–124 (1995).
30. Alfé, D. Price, G. D. & Gillan, M. J. Iron under Earth's core conditions: liquid-state thermodynamics and high-pressure melting curve from ab initio calculations. *Phys. Rev. B* **65,** 165118 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by an ERI Visiting Professorship to G.H. We thank A. Jackson for comments and O. Lord for experimental references. Data were provided by the ORFEUS Data Center, de Bilt, the Netherlands, the J-Array data centre, ERI, Tokyo, Japan, and by NIED, Tsukuba, Japan.

Author Contributions Both authors contributed equally to the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.H. (george.helffrich@bris.ac.uk).

METHODS

S*m*KS analysis. We investigate the family of SmKS ($m = 2, 3, 4$ and 5) waves using differential travel times between SmKS waves with different values of m. These are sensitive to the velocity structure of the uppermost few hundred kilometres of the outer core (see Fig. 2b for the bottoming depths of SmKS). S4KS and S5KS sample more predominantly the shallowest 100 km of the core, and their differential travel times are insensitive to the velocity structure of the mantle. Robust observations of S4KS and S5KS have been hampered by their mutual interference, which we ameliorate by array analyses of SmKS at large distances, from 140 to 170°, where the arrivals separate. We analyse broadband seismograms of two deep earthquakes in Fiji–Tonga observed in Europe and an event in Argentina observed in Japan and Taiwan (Supplementary Table 1). The selected events have quite simple impulsive source time functions, such that their SmKS phases are separated clearly enough from each other up to $m = 5$, as seen in the record sections of the radial component of the broadband seismograms (Fig. 1c and Supplementary Fig. 2). Comparisons with the record sections drawn for reflectivity synthetic seismograms[31] for the same event–network pairs (Fig. 1c and Supplementary Fig. 2) show systematic differences across the arrays in the SmKS differential travel times (dt) between the observations and PREM. In particular, S3KS (relative to S2KS) and S4KS and S5KS (both relative to S3KS) all arrive later than those computed for PREM and have about the same delay (~1 s). To quantify these observations accurately, differential travel times and slownesses are measured by array techniques using either SKKS (called S2KS hereafter, for clarity) or S3KS as the reference phase. Figure 2a shows differential travel times that are measured on the linearly stacked waveforms for the distances at the array centres and corresponding slownesses between three pairs of SmKS waves; S3KS relative to S2KS (called S3KS–S2KS, labelled $dt^{3,2}$), S4KS relative to S3KS (S4KS–S3KS, $dt^{4,3}$) and S5KS relative to S3KS (S5KS–S3KS, $dt^{5,3}$; see Supplementary Table 2 for data).

SmKS touches an internal caustic and reflects at the underside of the CMB $m - 1$ times, suffering a phase delay of nearly $\pi(m - 1)/2$ relative to SKS. Consequently, SmKS and S($m + 1$)KS are related by a Hilbert transform and SmKS and S($m + 2$)KS have opposite polarities. To account for the phase shift, we Hilbert-transform linearly stacked S2KS. We then slant-stack the observed broadband seismograms aligned on S2KS (Supplementary Table 2), and measure the arrival time difference between the transformed S2KS and S3KS. For S4KS–S3KS, S3KS is used as the alignment phase. The stacked S3KS is Hilbert-transformed and the relative travel time of S4KS is measured on the slant-stacked waveforms. S5KS–S3KS is measured as the arrival time difference between S3KS and a pulse of the opposite polarity on the slant-stacked trace aligned on S3KS (Supplementary Figs 2 and 3 and Supplementary Table 2). Synthetic tests on the reflectivity seismograms for several global Earth models show that these measurements give differential travel times between SmKS waves that agree with their ray theoretical predictions to within 0.2 s. Slight differences in instrument responses among the stations hardly affect the results of array processing according to a synthetic test, so stacking and dt measurements are performed on the broadband seismograms to avoid deconvolution (Supplementary Fig. 3). S3KS–S2KS and S4KS–S3KS are measured by picking peaks and cross-correlating the stacked waveforms, whereas S5KS–S3KS is measured by picking peaks only. The errors in those measurements are evaluated on the basis of the 95% confidence levels derived from stacking (Supplementary Fig. 3 and Supplementary Table 2).

Inversion methodology. The measured differential times and differential slownesses (dp) relative to those computed from PREM are listed in Supplementary Table 2 and shown in Fig. 2a, and the v_P anomaly as a function of depth is sought relative to PREM. The choice of PREM as the reference model is motivated by the many previous studies[10,11,32] which showed that PREM agrees with the SmKS differential travel times better than any other global reference model, such as IASP91, AK135 and SP6. Although previous studies[9–11] suggested the presence of a thin layer (50–100 km thick) with a v_P 1–2% lower than PREM at the top of the outer core, a recent study[12] shows that the deviations from PREM, if any, should be smaller.

We use a $\tau - p$ inversion method to invert the differential time and slowness measurements to determine the optimum velocity profile. A variety of basis functions were explored to eliminate parameterization bias. Fit details and coefficients are provided in Supplementary Table 3.

As a result of the inversion, we find that v_P values are slightly slower than predicted by PREM for the top 300 km of the outer core. As seen in Fig. 2a, b, the new data of S5KS–S3KS and S4KS–S3KS do not favour a steep and monotonic reduction of v_P as proposed in the previous studies, although a moderately lower v_P (by ~0.3%) than predicted by PREM at the CMB is certainly required. Near the CMB, v_P is consistent with the range of feasible solutions in ref. 12. According to synthetic tests, density perturbations up to about 1% from PREM do not affect the observed relative amplitudes of SmKS above the uncertainty level of the stacked seismograms, so we do not have a tight control over density, unlike that over v_P. Our model is consistent with the measurements of S4KS–S3KS obtained for individual seismograms[10] and for a small-aperture array[32], although the uncertainties in the previous measurements are far larger than ours. The variations in global compilations of S3KS–S2KS for individual station–earthquake pairs are much larger but are broadly consistent with our model, probably because they are prone to bias by mantle-side, small-scale velocity heterogeneity.

The synthetic seismograms are computed by the reflectivity method, with the stacked broadband seismograms of S2KS used for the source time function after correction of phase shift. We emphasize that data for both Fiji events suggest similar v_P anomalies and that the models (for example KHOCQ; Supplementary Fig. 4) obtained for the Fiji events match the differential travel times for the Argentina event. The D′′ region sampled by the data from Fiji to Europe includes a root of the Pacific super plume, and that for Argentina to Japan is near a subducted plate (Fig. 1a and Supplementary Fig. 1). The synthetic stacked seismograms for the derived model (KHOCQ) show excellent fits to the observed stacked seismograms for both the Fiji and the Argentina events (Supplementary Fig. 3). The presence of a layer with a large velocity anomaly in D′′ including an ultralow-velocity zone changes S3KS–S2KS, S4KS–S3KS and S5KS–S3KS by about 0.2 s, but their relative magnitudes are hardly affected (less than 0.1 s). The minor difference in the models obtained for the two regions may be partly explained by the difference in the v_S structure in D′′. More extreme and smaller-scale anomalies might affect the differential travel times, but piercing points at the CMB of the SmKS rays are spread over 3,000 km and 2,000 km beneath the receiver side and source side, respectively. The differential travel times used, therefore, should reflect the mantle and core structure averaged over a wide area. All of these considerations indicate that lateral variation in the mantle-side structure has only minor effects on the model, and that there actually is a velocity anomaly in the outermost core.

Core liquid model. The thermodynamic properties of dominantly iron liquids containing various impurities are well known owing to their commercial importance in steelmaking. One example is the Fe–O–S system. Experimental constraints are available on room-pressure thermophysical properties, leading to tabulations of their thermodynamic properties and models for mixing in the liquid state. For metallurgical and materials science reasons, the high-pressure properties of solids are also known through pressure–volume equation-of-state measurements.

Liquid properties are not as well characterized at high pressure, however. A thermodynamic approach is used to obtain the properties of Fe–O–S liquids from the melting curves of Fe, FeO and FeS, which have been determined experimentally[33,34], and shock-wave measurements on FeO[35]. At the melting point, the free energy, G, of the solid (s) and liquid (l) are equal. Referring to the standard states, G^0, of the solid and liquid at reference temperature and pressure (T_r, P_r), and using the dependence of free energy on entropy, S, and volume, V, we find that

$$G(P,T) = G^0(P_r, T_r) + \int_{P_r}^{P} V(p,T)\,\mathrm{d}p - \int_{T_r}^{T} S(P_r, t)\,\mathrm{d}t$$

$S(T)$ is available from standard tables at ambient pressure, P_r, and $V(P_r, T)$ may be obtained from thermal expansivity measurements on solids at ambient pressure yielding $\alpha(T)$, through the relation

$$V(T) = V(T_r)\exp\left[\int_{T_r}^{T}\alpha(t)\,\mathrm{d}t\right]$$

Solid volumetric properties are obtained by a third-order Birch–Murnaghan equation of state. The implicit relation between P and V is through a reference volume, V_0, from which the finite strain, f, undergone by compression is expressed as $f = [(V/V_0)^{2/3} - 1]/2$. The equation of state depends on isothermal bulk modulus, K_0, and its pressure derivative, K':

$$P(f) = 3K_0 f(1+2f)^{5/2}[1-3f(K'-4)/4] \quad (1)$$

The required integral, $\int V\,\mathrm{d}P$, is usually evaluated by integrating equation (1) by parts[36], yielding

$$\int_{P_r}^{P} V\,\mathrm{d}p = (V - V_r)(P - P_r) - \int_{V_r}^{V} P(\mathrm{d}f/\mathrm{d}v)\,\mathrm{d}v$$

In equation (1), K_0 is the bulk modulus at P_r but at high T. The Anderson–Grüneisen parameter, δ_T, is used to weaken the bulk modulus at high temperature, under the assumption that it is constant above the Debye temperature[37]. Thus, $K_0(T) = K_0(T_r)\exp\left[-\delta_T\int_{T_r}^{T}\alpha\,\mathrm{d}t\right]$. K' is taken to be independent of temperature, and, if not otherwise known, $K' \approx \delta_T$ (ref. 38).

To calculate $G_l(P, T)$, liquid K_0, $K' \equiv \delta_T$ and α are needed. These are estimated from the melting curves, where $G_l = G_s$, using solid equation-of-state data to calculate G_s and matching the experimental melting brackets. Grid search over K_0, α and K' yields the values shown in Supplementary Table 4, and Supplementary Fig. 5 shows the results of the fits to the experimental brackets for the liquids. For

higher-pressure calculations, entropies of the solid are modified for known phase transitions shown in Supplementary Fig. 5. These lead to minor changes in the calculated melting-curve slope but it is unresolvable given the width of the melting brackets. For this reason, owing to their minor effect on gross thermodynamic properties, structural changes in liquids are ignored[39]. Non-ideal mixing of end-member liquid is required to fit Fe–FeS eutectic composition variation with pressure; see Supplementary Information for justification and parameterization.

Velocities in the liquid are calculated from the adiabatic bulk modulus, K_s, and density, ρ, at high P and T using the relation $v_P = \sqrt{K_s/\rho}$. The adiabatic bulk modulus is obtained from the isothermal one using $K_s = K(1 + T\alpha\gamma)$. The core liquid Grüneisen parameter, γ, is 1.52 (ref. 30). The high-pressure thermal expansivity used here depends on the compressed volume[40] via $\mathrm{d}\ln(\alpha)/\mathrm{d}\ln(V) = \delta_T(V/V_0)^{\kappa}$, with $\kappa = 1.4$.

Calculation of the adiabatic gradient in the core, $\mathrm{d}T/\mathrm{d}z = g\alpha/C_p$ requires values for gravitational acceleration, g, and heat capacity, C_p. We calculate g at any depth in the core by linear interpolation between its value at the CMB given by PREM[13] and zero at the Earth's centre. The high-pressure heat capacity is obtained from $G_l(P, T)$ through the thermodynamic identity $C_p = -T(\partial^2 G/\partial T^2)$, evaluated numerically.

Adiabatic compression of a homogeneous core liquid. When a homogeneous liquid is adiabatically compressed, $\Phi(r) = v_P^2(r)$ varies approximately with radius as follows[1]:

$$1-\frac{1}{g}\frac{\mathrm{d}\Phi}{\mathrm{d}r}=\left(\frac{\partial K}{\partial P}\right)_T+\left[2\left(\frac{\partial K}{\partial P}\right)_T-1+\gamma-3\delta_T\right]T\alpha\gamma$$

Here $\delta_T = -(1/\alpha K)(\partial K/\partial T)_P$ and K, γ and α are the isothermal bulk modulus, the Grüneisen parameter and the thermal expansion coefficient, respectively. $K(P)$ is computed using the third-order Birch–Murnaghan equation. To demonstrate the smoothness of a profile of $1 - (1/g)\mathrm{d}\Phi/\mathrm{d}r$ for a compressed homogenous liquid, we present an example that fits PREM well (Fig. 2c). The following parameters are used for the computation: $K_0 = 100$ GPa, $K' = 4.75$, $\rho_0 = 6{,}280\ \mathrm{kg\,m^{-3}}$, $\gamma = 0.8$, $T_{\mathrm{CMB}} = 4{,}300$ K, $\mathrm{d}T/\mathrm{d}r = -0.3\ \mathrm{K\,km^{-1}}$ and $\delta_T = 3.33$. The thermal expansion coefficient, α, decreases linearly from $5\times10^{-6}\ \mathrm{K^{-1}}$ at the CMB to $3\times10^{-6}\ \mathrm{K^{-1}}$ at 1,700 km below the CMB.

31. Kind, R. Extensions of the reflectivity method for a buried source. *J. Geophys.* **45,** 373–380 (1979).
32. Eaton, D. W. & Kendall, J.-M. Improving seismic resolution of outermost core structure by multichannel analysis and deconvolution of broadband S*m*KS phases. *Phys. Earth Planet. Inter.* **155,** 104–119 (2006).
33. Boehler, R. Melting of the Fe-FeO and the Fe-FeS systems at high-pressure - constraints on core temperatures. *Earth Planet. Sci. Lett.* **111,** 217–227 (1992).
34. Boehler, R. Temperatures in the Earth's core from melting-point measurements of iron at high static pressures. *Nature* **363,** 534–536 (1993).
35. Svendsen, B., Anderson, W. W., Ahrens, T. J. & Bass, J. D. Ideal Fe-FeS, Fe-FeO phase relations and Earth's core. *Phys. Earth Planet. Inter.* **55,** 154–186 (1989).
36. Fei, Y., Saxena, S. K. & Navrotsky, A. Internally consistent thermodynamic data and equilibrium phase relations for compounds in the system MgO-SiO2 at high pressure and high temperature. *J. Geophys. Res.* **95,** 6915–6928 (1990).
37. Anderson, O. L., Isaak, D. & Oda, H. High-temperature elastic constant data on minerals relevant to geophysics. *Rev. Geophys.* **30,** 57–90 (1992).
38. Bina, C. R. & Helffrich, G. Calculation of elastic properties from thermodynamic equation of state principles. *Annu. Rev. Earth Planet. Sci.* **20,** 527–552 (1992).
39. Morard, G. *et al.* Structure of eutectic Fe-FeS melts to pressures up to 17 GPa: implications for planetary cores. *Earth Planet. Sci. Lett.* **263,** 128–139 (2007).
40. Anderson, O. L., Oda, H., Chopelas, A. & Isaak, D. A thermodynamic theory of the Grüneisen ratio at extreme conditions: MgO as an example. *Phys. Chem. Miner.* **19,** 369–380 (1993).

doi:10.1038/nature09634

Gene expression divergence recapitulates the developmental hourglass model

Alex T. Kalinka[1]*, Karolina M. Varga[1]*†, Dave T. Gerrard[2], Stephan Preibisch[1], David L. Corcoran[3], Julia Jarrells[1], Uwe Ohler[3], Casey M. Bergman[2] & Pavel Tomancak[1]

The observation that animal morphology tends to be conserved during the embryonic phylotypic period (a period of maximal similarity between the species within each animal phylum) led to the proposition that embryogenesis diverges more extensively early and late than in the middle, known as the hourglass model[1,2]. This pattern of conservation is thought to reflect a major constraint on the evolution of animal body plans[3]. Despite a wealth of morphological data confirming that there is often remarkable divergence in the early and late embryos of species from the same phylum[4–7], it is not yet known to what extent gene expression evolution, which has a central role in the elaboration of different animal forms[8,9], underpins the morphological hourglass pattern. Here we address this question using species-specific microarrays designed from six sequenced *Drosophila* species separated by up to 40 million years. We quantify divergence at different times during embryogenesis, and show that expression is maximally conserved during the arthropod phylotypic period. By fitting different evolutionary models to each gene, we show that at each time point more than 80% of genes fit best to models incorporating stabilizing selection, and that for genes whose evolutionarily optimal expression level is the same across all species, selective constraint is maximized during the phylotypic period. The genes that conform most to the hourglass pattern are involved in key developmental processes. These results indicate that natural selection acts to conserve patterns of gene expression during mid-embryogenesis, and provide a genome-wide insight into the molecular basis of the hourglass pattern of developmental evolution.

The notion that early development is similar among related animal species has been a guiding principle in comparative embryology since von Baer (1828) formalized the observation as his third law[10]. Darwin (1859) believed this to be the most compelling evidence in favour of common descent, reasoning that adult life-stages will afford the greatest opportunity for natural selection to operate, and thus adult structures should show signs of species-specific adaptations more than earlier stages[11]. These earlier stages, where adaptive opportunities are limited, will ultimately represent the 'pruned' but necessary features of ancestral differentiation[12].

Despite its intuitive appeal, the principle of early embryonic conservation has not been supported by morphological studies[2]. Counter to the expectations of early embryonic conservation, many studies have shown that there is often remarkable divergence between related species both early and late in development, often with little apparent influence on adult morphology[4–7]. The extensive variation that is seen in early and late development is contrasted by a period of conserved morphology occurring in mid-embryogenesis. This is known as the phylotypic period because it coincides with a period of maximal similarity between the species within each animal phylum[13].

The morphological conservation evident in the phylotypic period motivated a proposal of the hourglass model[1,2] as a revised formulation of von Baer's third law. The hourglass model predicts that early and late divergence is separated by a 'waist' corresponding to the phylotypic period. One of these studies argues that an increase in the number of global interactions between genes and developmental processes during the phylotypic period renders any evolutionary modification highly deleterious due to their damaging side-effects[2], whereas the other study views conservation during this period as a consequence of the need for precise coordination between growth and patterning, which is seen to be reflected in the genomic organization of the vertebrate Hox genes[1].

Support for the hourglass model has been found at the morphological[7,14] and sequence levels[15–17]. However, both the model and the concept of the phylotypic period remain controversial subjects in the literature[3,18], with some studies of heterochrony in vertebrates indicating that divergence peaks at the phylotypic period[19] or that there is no temporal pattern of phenotypic conservation[20].

Although it is generally appreciated that gene expression divergence has a key role in the evolution of morphological diversity[8,9], no studies so far have addressed the extent to which expression divergence underpins the morphological hourglass pattern at the genome-wide level. Here, we test the molecular basis of the hourglass model of developmental evolution using gene expression data from six *Drosophila* species with sequenced genomes (*D. melanogaster*, *D. simulans*, *D. ananassae*, *D. persimilis*, *D. pseudoobscura* and *D. virilis*), thereby enabling unambiguous quantitative comparisons across orthologous genes for a set of species separated by up to 40 million years. Gene expression levels were measured for 3,019 genes, known to be expressed during embryonic development from RNA *in situ* data[21], at 2-h intervals for the majority of embryogenesis using a microarray time course with three biological replicates per species and four species-specific probes per gene (Supplementary Figs 1 and 2).

For each gene in each species we generated a gene expression time course, corrected for differences in developmental time (Supplementary Information, Section 2.2), and measured the correlation of the resulting temporal profiles for each pair of species (Fig. 1a, b). The distribution of the correlation coefficients shows that whereas most genes are positively correlated in their temporal expression, the divergence in embryonic gene expression follows the known phylogenetic relationships[22]. These results clearly demonstrate that there is evolutionary signal across the data set as a whole.

To quantify gene expression divergence rigorously we fitted a linear model to the expression data. This approach enables us to quantify the divergence between species by measuring the influence that different species have on the expression of individual genes at specific times during development. We extract two different measures of divergence from the model: quantitative divergence, which reflects differences in expression across the whole time course; and temporal divergence, which reflects divergence of temporal profiles at specific time points (Supplementary Information, Section 2.6). We show that both of these measures of

[1]Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, 01307 Dresden, Germany. [2]Faculty of Life Sciences, The University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK. [3]Institute for Genome Sciences and Policy, Duke University, 101 Science Drive, Durham NC 27708, USA. †Present address: New Biochemistry Building, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK.
*These authors contributed equally to this work.

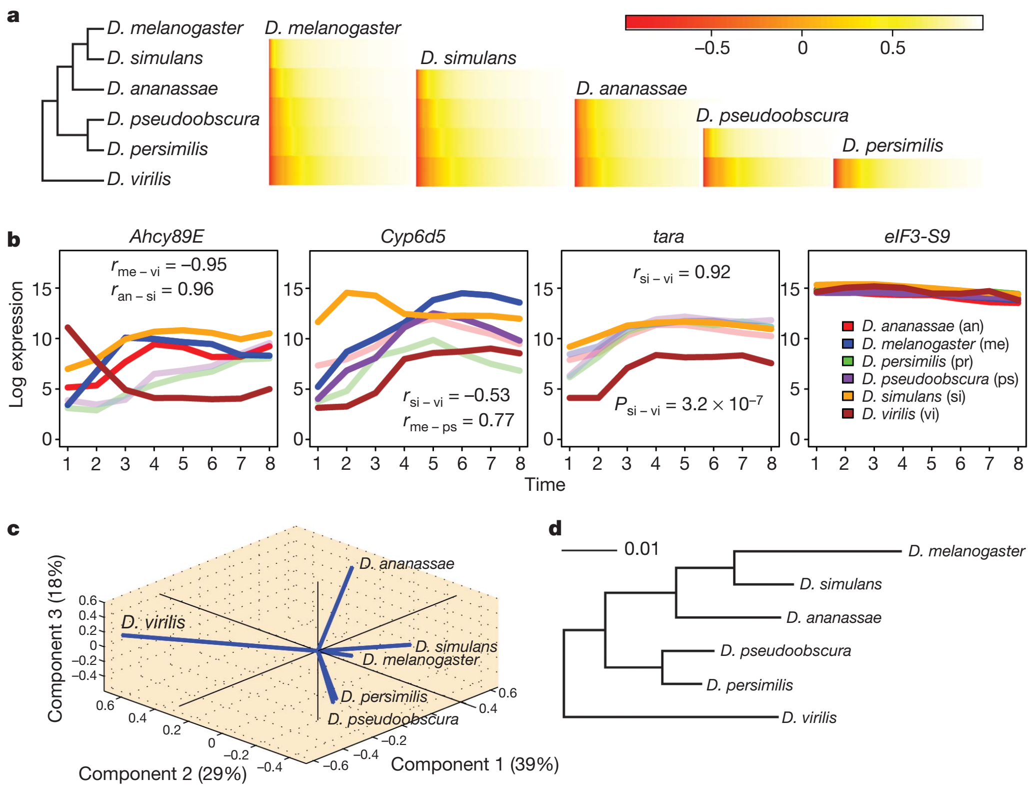


Figure 1 | Gene expression during *Drosophila* embryogenesis recapitulates the known phylogeny. a, Between-species pairwise correlation coefficients for temporal profiles are depicted using a colour gradient. **b**, Species profiles are shown for two genes with both positive and negative correlations between different species pairs (*Ahcy89E* and *Cyp6d5*) and two genes that are temporally conserved (*tara* and *eIF3-S9*). Log_2 expression profiles are averaged over probes and replicates. Selected correlation coefficients are shown on the plots, and the *P*-value refers to quantitative divergence. Time points along the *x* axis are 2-h intervals starting from 0–2 h (1) and ending at 14–16 h (8). **c**, The first three principal components for quantitative divergence. **d**, A maximum likelihood phylogeny based on temporal expression divergence across all genes.

divergence recapitulate the known evolutionary relationships between the species when the phylogeny is constructed using all of the genes simultaneously (Fig. 1c, d). However, despite producing an identical topology to the known phylogeny, we see relatively long terminal branches in the phylogram, indicating that gene expression divergence does not scale with the amount of time separating pairs of species (Supplementary Fig. 3)[23].

If temporal expression divergence has saturated through time then we would expect to find a reduced capacity for reconstructing the known phylogeny at the level of individual genes. To explore this possibility we estimated the phylogenetic signal for each gene using a statistic that compares the observed phylogenetic signal to what would be expected under a process of random evolutionary change[24]. A random evolutionary process produces a phylogeny where closely related species resemble each other more than distantly related species as lineages inherit the random changes of their ancestors. The results show that at each time point the majority of genes exhibit a weaker phylogenetic signal than expected under random evolution (Supplementary Fig. 5a).

Phylogenetic signal may be eroded by stabilizing selection[23], and to test for this possibility we compared different evolutionary scenarios by fitting four alternative models to the expression data. The models were purely random evolutionary change and three stabilizing selection models where the optimal expression level may vary between groups of species, allowing us to model adaptive changes in expression (Supplementary Fig. 4)[25]. The stabilizing selection models describe the change in expression as a combination of random changes and stabilizing selection curtailing the accumulating variance. The results show that at each time point at least 80% of genes fit best to models that incorporate stabilizing selection (Supplementary Fig. 5b). We also see that a substantial fraction of the genes fit best to models where there are adaptive changes in expression, indicating that a combination of both stabilizing and directional selection may be acting on a large fraction of the genes[26].

The variance between species in the behaviour of a particular gene at a particular time point provides a measure of the divergence of a gene's temporal dynamics (see Methods). Plotting these values across all genes as a function of time shows that temporal expression divergence follows an hourglass pattern with maximal conservation occurring at time point 5 (8–10 h), a period that corresponds to the extended germband stage, generally regarded as the arthropod phylotypic period (Fig. 2a). We confirmed that the hourglass pattern is not an aggregate behaviour of the data set, but is present on a gene-by-gene basis for the majority of genes (Supplementary Figs 6 and 7), and also that this pattern is evident in the absolute, untransformed gene expression levels (Supplementary Fig. 8b and Supplementary Information, Sections 2.6 and 2.7). For genes that fit best to models where the optimal expression level is the same across all species we calculate a measure of selective constraint[23] (Fig. 2b). This shows that for genes whose evolutionary optimum is the same across species, selective constraint is maximized during the phylotypic period when gene expression divergence is minimized. Therefore, natural selection conserves gene expression patterns during the phylotypic period.

To discover the functional classes of genes responsible for driving the hourglass pattern in the data, we correlated each gene's divergence profile with the average across all genes, thereby allowing us to rank genes by their tendency to follow the global hourglass divergence profile. We find that these genes are enriched for biological processes involved in cellular and organismal development and gene expression (Supplementary Tables 1–3). Moreover, functional characterization of genes that follow an absolute expression hourglass (Supplementary Fig. 8b) shows that they are also enriched for developmental and gene expression processes (Fig. 3a and Supplementary Tables 4–6). Taken together, these results show that genes involved in core developmental processes conform

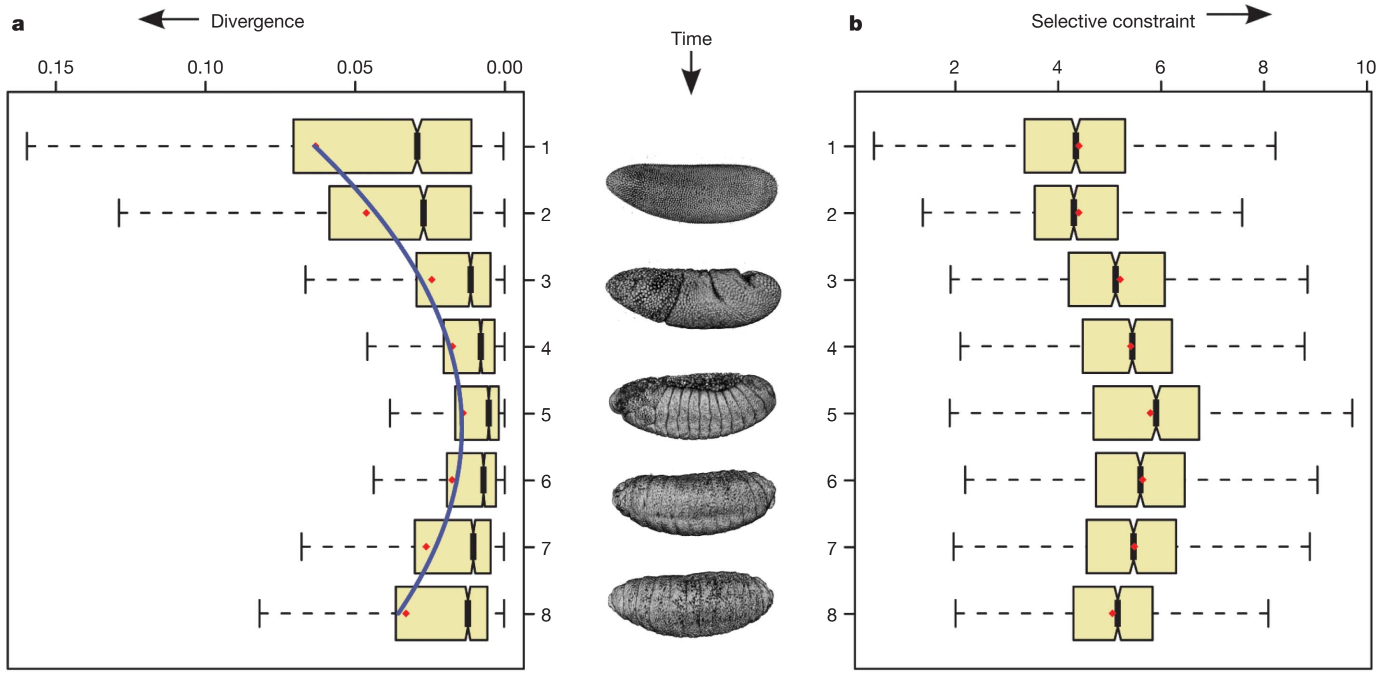


Figure 2 | Temporal expression divergence is minimized during the phylotypic period. **a**, Temporal divergence of gene expression at individual time points during embryogenesis. The curve is a second-order polynomial that fits best to the divergence data. Embryo images are three-dimensional renderings of time-lapse embryonic development of *D. melanogaster* using Selective Plane Illumination Microscopy (SPIM). **b**, Selective constraint for genes that fit best to single optimum stabilizing selection models, calculated as the negative log of the equilibrium variance (see Methods and Supplementary Fig. 5b). Time points are 2-h intervals starting from 0–2 h (1) and ending at 14–16 h (8). Red diamonds indicate the mean; error bars encompass data within 1.5 times the inter-quartile range, and the boxes show the lower and upper quartiles together with the median.

strongly to the global hourglass divergence pattern both in terms of temporal dynamics and absolute expression differences.

We also asked whether there are sets of genes that don't follow the global hourglass pattern and found genes enriched for processes involved in secondary metabolism, the immune system, and responses to oxidative and wounding stresses (Supplementary Fig. 9a and Supplementary Table 7). These are processes that are upregulated late in development, such as pigment or chitin metabolism, or processes that will be upregulated in response to changes that are independent of the developmental program, such as a change in the external environment or the presence of a parasite. The transcript levels of genes in this latter category will reflect the particular challenges faced by individual embryos and so we would not expect these genes to follow a clear temporal pattern of conservation and divergence. These genes tend to be zygotically expressed and are largely present in the yolk (Supplementary Table 7).

Independent of the hourglass patterns, our measures of quantitative and temporal expression divergence exhibit similar functional associations; housekeeping processes tend to be conserved and metabolic processes tend to be divergent (Supplementary Tables 8–11 and Supplementary Information, Sections 2.8 and 2.9). Given these broad functional similarities, it is of interest to ask whether genes in these categories of divergence also share similar genomic and gene-level features. We observe that genes that diverge quantitatively tend to have short introns and 5′ intergenic regions (Fig. 3b) whereas genes that diverge temporally have long introns and 5′ intergenic regions consistent with the notion that increased regulatory complexity in long noncoding regions[27] may provide opportunities for temporal expression divergence (Fig. 3c). This increased regulatory complexity is also supported by a strong positive correlation between temporal divergence and tissue specificity. Additionally, temporal divergence is negatively correlated with mRNA length, raising the possibility that the

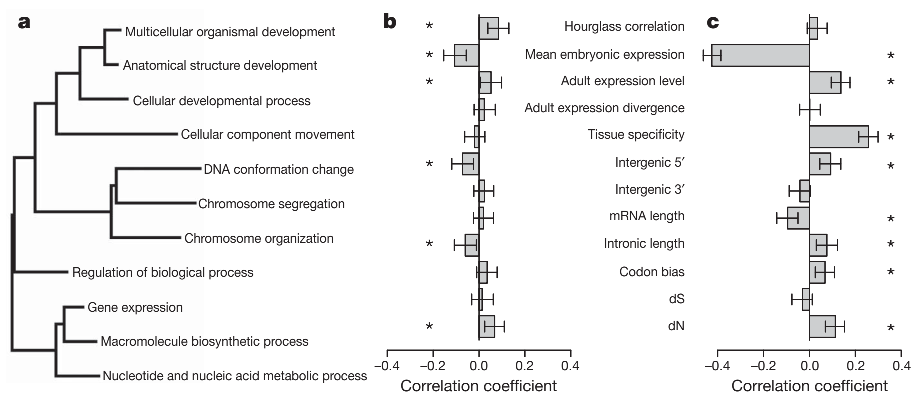


Figure 3 | Properties of genes with different divergence patterns. **a**, A neighbour-joining dendrogram of enriched functional processes for genes that follow an hourglass pattern of divergence. **b**, **c**, Correlation of gene-level variables with quantitative divergence (**b**) and temporal divergence (**c**). Error bars are 95% confidence intervals based on 1,000 bootstraps. Asterisks indicate significant correlations. dN, non-synonymous substitution rate; dS, synonymous substitution rate.

proteins of these shorter genes are engaged in fewer protein–protein interactions. We also observe a positive correlation between rates of amino acid evolution (dN) and both quantitative and temporal divergence, supporting similar findings based on adult expression levels[28], and providing further evidence that embryonic expression divergence is measuring biologically relevant signals.

Our results show that gene expression is more resistant to evolutionary change during mid-embryogenesis than either early or late periods of *Drosophila* development. Evolutionary analyses support the notion that this conservation is the result of natural selection acting to maintain expression levels and their temporal relationships during mid-embryogenesis for genes involved in building up the body plan of the larva. These results complement a recent finding suggesting that the pupal stage in *Drosophila* is under strong stabilizing selection due to the complexity of the processes that occur during metamorphosis, a process that parallels many aspects of embryonic development[29]. These findings seem to support the hypothesis of ref. 2 that an increase in global interactions constrains evolutionary change of the phylotypic period; however, neither study directly addresses the coordination of growth and patterning proposed by ref. 1. Such a relationship may be best examined in the context of gene regulatory networks. Future studies will also need to address the mode and strength of selection acting on gene expression with greater resolution by coupling interspecific expression divergence with intraspecific variation during embryogenesis[30].

METHODS SUMMARY

RNA was extracted from embryos from six *Drosophila* species (*D. melanogaster*, *D. simulans*, *D. ananassae*, *D. persimilis*, *D. pseudoobscura* and *D. virilis*) reared at 25 °C. The embryos were aged at 2-h intervals to form a time course. Sixty-base-pair-long, species-specific microarray probes (four per species) were selected by choosing regions of the orthologous genes of each species that were maximally conserved according to an information entropy measure. Candidate probes with a G+C content higher than 50% were penalized and hence were less likely to be chosen. After scaling the time courses and normalizing replicates, the following linear model was fitted to log expression levels:

$$\log(y_{ijklmn}) = \mu + G_i + S_j + T_k + r_{l(j)} + p_{m(ij)} + GS_{ij} + GT_{ik} + ST_{jk} + rG_{l(j)i} + pT_{mk(ij)} + rp_{l(j)m(ij)} + rT_{kl(j)} + GST_{ijk} + e_{n(ijklm)}$$

where μ is the global average, G_i is the gene effect, S_j is the species effect, T_k is the time effect, $r_{l(j)}$ is the replicate effect nested in species, $p_{m(ij)}$ is the probe effect nested in genes and species, and $e_{n(ijklm)}$ is the residual error. Values are averaged over missing subscripts. Divergence per time point was measured as the between-species variance in *GST* values for each gene separately. We fitted four different evolutionary models to the *GST* values for each gene using the R package 'ouch' and ranked them by their Akaike Information Criterion (AIC). The models were Brownian motion plus three stabilizing selection models with between one and three selective optima (Supplementary Fig. 4). Genomic features of genes were retrieved from FlyBase release 5.14, adult expression level and tissue specificity were retrieved from FlyAtlas, and tissue expression data were retrieved from APOGEE (http://fruitfly.org/cgi-bin/ex/insitu.pl). Partial correlations were calculated and 95% confidence intervals for each partial correlation were generated from 1,000 bootstraps.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 September; accepted 2 November 2010.

1. Duboule, D. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Dev. Suppl.* 135–142 (1994).
2. Raff, R. A. *The Shape of Life: Genes, Development and the Evolution of Animal Form* (Univ Chicago Press, 1996).
3. Hall, B. K. Phylotypic stage or phantom, is there a highly conserved embryonic stage in vertebrates? *Trends Ecol. Evol.* **12,** 461–463 (1997).
4. Sander, K. Specification of the basic body plan in insect embryogenesis. *Adv. Insect Physiol.* **12,** 125–238 (1976).
5. Wray, G. A. & Raff, R. A. Rapid evolution of gastrulation mechanisms in a sea urchin with lecithotrophic larvae. *Evolution* **45,** 1741–1750 (1991).
6. Goldstein, B., Frisse, L. M. & Thomas, W. K. Embryonic axis specification in nematodes: evolution of the first step in development. *Curr. Biol.* **8,** 157–160 (1998).
7. Schmidt, K. & Starck, J. M. Developmental variability during early embryonic development of zebra fish, *Danio rerio*. *J. Exp. Zoolog. B* **302,** 446–457 (2004).
8. Carroll, S. B. Evolution at two levels: on genes and form. *PLoS Biol.* **3,** e245 (2005).
9. Wray, G. A. The evolutionary significance of *cis*-regulatory mutations. *Nature Rev. Genet.* **8,** 206–216 (2007).
10. von Baer, K. E. *Uber Entwickelungsgeschichte der Thiere: Beobachtung und Reflektion* (Königsberg, 1828).
11. Darwin, C. *On the Origin of Species* (Murray, 1859).
12. Garstang, W. The theory of recapitulation: a critical restatement of the biogenetic law. *Linn. J. Zool* **35,** 81–101 (1922).
13. Sander, K. The evolution of patterning mechanisms: gleanings from insect embryogenesis and spermatogenesis. In *Development and Evolution* 137–159C (Cambridge Univ. Press, 1983).
14. Galis, F. & Metz, J. A. Testing the vulnerability of the phylotypic stage: on modularity and evolutionary conservation. *J. Exp. Zool.* **291,** 195–204 (2001).
15. Hazkani-Covo, E., Wool, D. & Graur, D. In search of the vertebrate phylotypic stage: a molecular examination of the developmental hourglass model and von Baer's third law. *J. Exp. Zoolog. B* **304,** 150–158 (2005).
16. Davis, J. C., Brandman, O. & Petrov, D. A. Protein evolution in the context of *Drosophila* development. *J. Mol. Evol.* **60,** 774–785 (2005).
17. Cruickshank, T. & Wade, M. J. Microevolutionary support for a developmental hourglass: gene expression patterns shape sequence variation and divergence in *Drosophila. Evol. Dev.* **10,** 583–590 (2008).
18. Richardson, M. K. *et al.* There is no highly conserved embryonic stage in the vertebrates: implications for current theories of evolution and development. *Anat. Embryol. (Berl.)* **196,** 91–106 (1997).
19. Bininda-Emonds, O. R. P., Jeffery, J. E. & Richardson, M. K. Inverting the hourglass: quantitative evidence against the phylotypic stage in vertebrate development. *Proc. R. Soc. Lond. B* **270,** 341–346 (2003).
20. Poe, S. & Wake, M. H. Quantitative tests of general models for the evolution of development. *Am. Nat.* **164,** 415–422 (2004).
21. Tomancak, P. *et al.* Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* **8,** R145 (2007).
22. Zhang, Y., Sturgill, D., Parisi, M., Kumar, S. & Oliver, B. Constraint and turnover in sex-biased gene expression in the genus *Drosophila. Nature* **450,** 233–237 (2007).
23. Bedford, T. & Hartl, D. L. Optimization of gene expression by natural selection. *Proc. Natl Acad. Sci. USA* **106,** 1133–1138 (2009).
24. Blomberg, S. P., Garland, T. & Ives, A. R. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57,** 717–745 (2003).
25. Butler, M. A. & King, A. A. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am. Nat.* **164,** 683–695 (2004).
26. Lemos, B., Meiklejohn, C. D., Cceres, M. & Hartl, D. L. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution* **59,** 126–137 (2005).
27. Nelson, C. E., Hersh, B. M. & Carroll, S. B. The regulatory content of intergenic DNA shapes genome architecture. *Genome Biol.* **5,** R25 (2004).
28. Nuzhdin, S. V., Wayne, M. L., Harmon, K. L. & McIntyre, L. M. Common pattern of evolution of gene expression level and protein sequence in *Drosophila. Mol. Biol. Evol.* **21,** 1308–1317 (2004).
29. Artieri, C. G. & Singh, R. S. Molecular evidence for increased regulatory conservation during metamorphosis, and against deleterious cascading effects of hybrid breakdown in *Drosophila. BMC Biol.* **8,** 26 (2010).
30. Rifkin, S. A., Kim, J. & White, K. P. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nature Genet.* **33,** 138–144 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Alexa for providing modified code for his topGO R package, A. Larracuente and T. Sackton for sharing data with us, M. Weber for generating the embryo images for Fig. 2, and Carl Zeiss MicroImaging for providing the SPIM microscope. We also thank N. Barton, T. Bedford, D. Hartl, J. Howard, A. Oates and D. Robertson for providing useful comments and discussion on the manuscript. This work was funded by The Human Frontier Science Program (HFSP) Young Investigator's Grant RGY0084.

Author Contributions K.M.V. and P.T. conceived the experiment, and K.M.V. and J.J. carried it out. K.M.V., P.T. and S.P. designed the microarray. P.T. conducted the interspecies correlation analysis, and S.P. formulated the linear interpolation algorithm. A.T.K. conceived and conducted the statistical analyses. D.T.G. and C.M.B. conducted the genomic correlates analysis. D.L.C. and U.O. carried out the probe orthology assignments. C.M.B. brought the hourglass concept to the attention of the HFSP team. A.T.K. wrote the paper with support from co-authors.

Author Information The expression data are available for download from ArrayExpress under experiment name 'hourglass', accession number E-MTAB-404, and together with the probe sequences at http://publications.mpi-cbg.de/4240-data. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.T. (tomancak@mpi-cbg.de).

METHODS

Embryo collections and RNA isolation and labelling. Embryos were collected from a population of well-fed adults reared at 25 °C. To synchronize the age of the embryos in each sample we pre-laid the flies twice for 1 h with a fresh apple juice plate with yeast paste before every collection. Another fresh plate with yeast was used to collect embryos. The plate was removed from the cage after a 2-h interval and aged in the same incubator for the remaining time required by each time point. After ageing, embryos were collected and rinsed with water to remove yeast paste, and then dechorionated in 100% bleach for 2 min and then washed in desalinated water. The embryos were then transferred into a 1.5-ml tube and snap-frozen in liquid nitrogen and stored at −80 °C.

When isolating RNA, embryos were thawed on ice and homogenized with a pellet pestle and a pellet pestle cordless motor (Kontes). RNA was isolated with the RNeasy Mini kit (Qiagen) and eluted with 30 μl of water. The RNA concentration was measured with the NanoDrop spectrophotometer and RNA quality was assessed with Bioanalyser using the Agilent RNA 6000 Nano kit.

To prepare samples for hybridization to the chip, we followed the Agilent One-Colour Microarray-Based Gene Expression Analysis protocol version 5.5. The starting amount of RNA was normalized to 600 ng for all samples. Samples of a given time-course were processed on the same day.

Probe selection. Probe selection was limited to 60-mers that started within 1 kb from the 3′ end of the transcript. The two main factors that influenced subsequent probe selection were the similarity of orthologous probes in six species determined by information entropy (Supplementary Fig. 1) and the specificity of a probe estimated by the G+C content-weighted BLAST score (Supplementary Information, Section 1.1).

Additionally, we incorporated the distance from the 3′ end of the transcript into the information entropy measure by means of weighting the information entropy (Supplementary Fig. 2). Hence, the further away the candidate probe was from the end of the transcript, the higher the final information entropy measure became.

Probe specificity was verified in two steps. We first rejected candidate probes that did not have a 60-nucleotide-long match to the respective genome assembly. By doing so we eliminated probes that fell on the border of two exons in the transcript sequence. For the remaining candidate probes the G+C content-weighted BLAST score was calculated. This score was the sum of nucleotides that were identical to the query 60-mer and were found in short, unspecific hits. The sum was weighted by the G+C content of hits shorter than 60 nucleotides. If the G+C content exceeded 50% the sum was multiplied by a factor greater than 1 and as a consequence the probe was penalized. Four probes were selected for each gene in each species and we calculated the base-pair overlap between the probes and, where possible, tried to minimize this value (Supplementary Fig. 13).

Time-course registration and correlation analysis. To register the time courses for different species with different developmental time periods[31] onto a common time axis we scaled the non-*melanogaster* time courses to *D. melanogaster* by maximizing the similarity among the profiles across all genes. The selection of genes on the array resulted in a progressive shift of signal intensity distributions from bimodal (mixture of non-expressed and expressed genes early) to unimodal (most genes expressed late) across the time course (Supplementary Fig. 18). Therefore we normalized the replicates for each time point in each time course separately using quantile normalization. Next we averaged the probe signal intensities using the Tukey biweight algorithm to obtain a single expression value per gene and time point in each species while removing outliers. We then re-sampled each time course to 100 time points using cosine transform interpolation (DCT)[32]. Subsequently, all 3,019 expression profiles of the non-*melanogaster* species were scaled by factors ranging from 0.4 to 1.6 in 0.01 increments to find the optimal scaling factor. We calculated squared sums of average differences between all the scaled profiles and *D. melanogaster* profiles and plotted these sums as a function of the scaling factor applied (Supplementary Fig. 16). The global minimum in the graph corresponded to the scaling factor at which all the profiles of the two species were most similar to each other. We applied the optimal scaling factors (Supplementary Table 12) to the averaged non-*melanogaster* profiles with the DCT interpolation resulting in registered time courses (four example genes, before and after registration, are shown in Supplementary Fig. 17).

To compare the overall shape of the profiles among species we row normalized the gene expression values for each gene in each time course and calculated pair-wise correlation coefficients for all pairs of orthologous genes. Genes ordered by this simple measure of similarity give an intuitive impression of the amount of conservation of temporal profiles among each pair of species (Fig. 1a).

For the statistical analysis described below, we applied the optimal scaling factors to the raw $\log_2$ Agilent array signal intensities and subsequently quantile-normalized each time point in each time course separately, as described above.

Linear models. A global ANOVA model was fitted to the data to partition the main effect variables and interactions of biological interest from random factors and residual error. This normalizes the gene expression values and provides a single coherent statistical framework in which to explore the variance and covariance structure of the data[33]. The model for gene expression, y_{ijklmn}, is a five-factor, partially nested, mixed-model ANOVA

$$\log(y_{ijklmn}) = \mu + G_i + S_j + T_k + r_{l(j)} + p_{m(ij)} + GS_{ij} + GT_{ik} + ST_{jk} + rG_{l(j)i} + pT_{mk(ij)} + rp_{l(j)m(ij)} + rT_{kl(j)} + GST_{ijk} + e_{n(ijklm)}$$

where μ is the global average, G_i is the gene effect, S_j is the species effect, T_k is the time effect, $r_{l(j)}$ is the replicate effect nested in species, $p_{m(ij)}$ is the probe effect nested in genes and species, and $e_{n(ijklm)}$ is the residual error. Values are averaged over missing subscripts. The probe and replicate effects are random factors in the model and account for error variance arising from different probes and from different samples of within-strain genotypes respectively.

The remaining terms are two- and three-way interactions between the main factors. The gene-by-species, $GS_{ij} = \log(y_{ij}) - G_i - S_j - \mu$, and gene-by-species-by-time, $GST_{ijk} = \log(y_{ijk}) - G_i - S_j - T_k - GS_{ij} - GT_{ik} - ST_{jk} - \mu$, effects contain information about divergence between species. Here we treat time as a categorical variable so that we can extract variances at discrete time points in different species. Divergence at each time point is then measured as the between-species variance in *GST* values per gene and per time point (Fig. 2a). Mean sums of squares were estimated for each variable in the model after subtracting the mean from the data, and the resulting ANOVA table is shown in Supplementary Information, Section 2.4. A Principal Component Analysis (PCA) of the gene × time (*GT*) effect from the above model was computed and the results are shown in Supplementary Information, Section 2.3.

A reduced version of the above model was fitted as a linear regression to each gene separately (the gene effect was dropped) using the R package 'limma' version 3.2.2 (ref. 34). Limma uses an empirical Bayesian approach to infer differential expression in individual genes, producing moderated *t*-statistics with Bayesian-adjusted denominators that incorporate information across the entire ensemble of genes[35]. By fitting a linear model to each gene separately, limma allows for gene-specific error distributions. The probe effect was also dropped from the ANOVA model as the probes were normalized using Tukey's median polish method to fit a linear model for gene expression to each gene, $y_{ij} = \exp_i + a_j + e_{ij}$, where $\exp_i$ is the normalized gene expression value for gene i, a_j is the probe effect for the jth probe, and e_{ij} is the residual error[36]. The species effect from limma is equivalent to the *GS* effect from the global ANOVA, and this value was used for assessing quantitative divergence between species (Fig. 3b).

The temporal profiles of genes were compared across species using a PCA-based approach[37]. This method quantifies pairwise species differences in temporal profiles for individual genes using the Mahalanobis distance, which is calculated using *GST* values for all time points estimated from the global ANOVA model. The Mahalanobis distance is calculated as

$$D_i^2 = (\Delta Z_i - Z_C)\mathrm{cov}(\Delta Z)^{-1}(\Delta Z_i - Z_C)^T,$$

where ΔZ_i is the species *GST* score contrast for gene i, Z_C is the centroid for all of the *GST* score contrasts, and cov(ΔZ) is the covariance matrix for the difference matrix ΔZ. This metric is distributed according to a chi-squared distribution with k degrees of freedom where k is the number of principal components included in the contrast. We used the Mahalanobis distances as a measure of temporal divergence between species across all time points (Fig. 3c). The distances were calculated using the first three principal components, which together account for 89% of the total variance.

Phylogenetic analyses and evolutionary models. A maximum likelihood phylogeny was constructed with *GST* values from every gene and every time point using the 'contml' continuous character restricted maximum likelihood approach implemented in PHYLIP version 3.69 (ref. 38) with *D. virilis* identified as the outgroup, and the resulting phylogram was plotted using Dendroscope version 2.0 (ref. 39) (Fig. 1d).

We estimated the phylogenetic signal for the *GST* values at each gene by calculating the *K* statistic described in ref. 24 using the R package 'picante' version 1.1-1. The tree used for this purpose was a phylogram based on median dS values for ∼10, 000 orthologous genes[40] which was then converted to a chronogram in the R package 'ape' version 2.5-1[41].

Ornstein–Uhlenbeck (OU) and Brownian motion models were fitted to the six species-specific *GST* values for each gene at each time point using the R package 'ouch' version 2.6-1 (ref. 25). The OU models fitted to each gene describe evolutionary change in a trait X over an infinitesimally small increment of time as $dX(t) = \alpha(\theta - X(t))dt + \sigma dB(t)$ where $dB(t)$ describes Brownian motion (independent and identically distributed normal random variables with mean 0 and variance dt), σ is the strength of Brownian motion, α is the strength of stabilizing selection, and θ is the trait optimum[42]. Under a purely Brownian process of

evolutionary change the first term on the right-hand side is absent. This model was extended by ref. 25 to include branch-specific values for θ, thereby allowing for adaptive evolution along specific branches. We fitted four models to each gene: Brownian motion plus three OU models with between one and three stabilizing selection optima (Supplementary Fig. 4), based on the chronogram mentioned above. Here we did not engage in an exhaustive model fitting endeavour as our intention was to demonstrate two things: (1) models incorporating stabilizing selection fit best to the majority of the genes, and (2) models incorporating adaptive changes in trait optima often out-perform non-adaptive models.

To avoid treating time points as if they are independent we fitted models to subsets of time points. These subsets were chosen by bootstrapping (1,000 bootstraps) hierarchical clusters of time points based on *GST* values for each gene separately using the R package 'pvclust' version 1.2-1 (ref. 43) and selecting clusters of time points with *P*-values below 0.05. This approach allows different modes of selection to operate across different periods of each gene's time course.

For each gene the model that showed the best fit to the data was defined as the model with the lowest Akaike Information Criterion (AIC), calculated as $2k - 2\log(\text{likelihood})$ where k is the number of degrees of freedom in the model in question. AIC scores balance the likelihood of a model against its complexity (the number of parameters in the model).

After ranking models by their AIC scores, genes for which Brownian motion was not ranked first were tested to see if the top-ranked model showed a significantly better fit to the data than Brownian motion using a log-likelihood ratio test. The resulting *P*-values were adjusted using the Benjamini–Hochberg false discovery rate correction in the R package 'multtest' version 2.4.0 (ref. 44) and models with adjusted *P*-values above 0.05 were dropped down into the Brownian category. We then repeated this process, but treating single optimum models as the null model and testing models that ranked best with two or three optima against this null model to ensure that the resemblance to the phylogeny for these genes was not the result of chance under a single optimum across all species. If single optimum models showed a better fit then they were, in turn, tested against the Brownian model.

We extracted a measure of selective constraint from the genes that fitted best to single optimum models (Fig. 2b), calculated as the negative log of the equilibrium variance, $\frac{\sigma^2}{2\alpha}$ (ref. 23).

Gene Ontology and tissue expression enrichment. Gene Ontology (GO) analyses were conducted using the R package 'topGO' version 1.14.0 (ref. 45). Three enrichment methods were used. For genes that were ranked by a real number score (such as a correlation coefficient) a Kolmogorov–Smirnov ranking test was applied and GO terms with distributions among the genes that showed significant departure from a uniform distribution in a particular direction were deemed to be enriched. Unranked sets of genes were tested for enriched GO terms using the 'elim' and 'parent-child' algorithms in topGO. The 'elim' algorithm decorrelates the local GO graph structure to take into account local dependencies between terms so that more biologically relevant terms are enriched[45] and the 'parent-child' algorithm controls for the inheritance bias between parent and child terms in the GO hierarchy[46]. Fisher's exact test was then used to determine enrichment *P*-values for both of these algorithms. The same approach was used to identify enriched tissue expression terms from a controlled vocabulary based on *in situ* expression data[21] by using modified code from the topGO package.

P-values from Kolmogorov–Smirnov tests were adjusted using the Benjamini–Hochberg false discovery rate correction, but no correction was applied to the 'elim' and 'parent-child' *P*-values because they are not calculated independently for each GO term in these algorithms and are effectively already adjusted. For defined sets of genes, the reference set was all of the genes on the chip.

We plotted a neighbour-joining tree of enriched, non-redundant GO terms[47] for Fig. 3a using the R package 'ape' and Dendroscope. Terms were enriched for 1,188 genes that show an hourglass profile in both temporal dynamics and absolute expression levels using Fisher's exact test and selecting terms with adjusted *P*-values below 0.05.

Correlation of divergence with gene-level variables. Quantitative and temporal divergence measures were generated for each of the 15 pairwise species comparisons. Following ref. 48, we converted these to nine branch lengths on the known phylogeny using the Fitch–Margoliash least squares method (implemented in the PHYLIP program 'fitch'[38]). Negative branch lengths were set to zero. Total expression divergence for each gene is the sum of branch lengths and constitutes our 'quantitative' and 'temporal' measures using the limma or Mahalanobis distances, respectively.

We collated structural, functional and expression data for all of the genes on the chip from public databases and previous genome-level studies. These data were generated from gene coordinates retrieved from FlyBase Release 5.14 (January 2009). Only protein-coding genes were retained (as all genes on our chip are protein coding) and genes from the heterochromatic portions of the otherwise 'euchromatic' chromosome arms were discarded (168 genes from the genome, including 25 from our chip data set)[49].

In addition, data on further variables for 8,500 *D. melanogaster* genes compiled by ref. 48 were obtained from the authors. These data could be assigned to 2,526 of the genes on the chip. Gene expression was described by adult expression level and tissue specificity (both from FlyAtlas[50]), expression divergence between adults (measured in a very similar set of species by ref. 22), and we added the mean embryonic expression level from our own data. Gene sequence evolution was described by codon bias (the frequency of optimal codons) in *D. melanogaster* and by dN and dS, the rates of non-synonymous and synonymous nucleotide substitutions, respectively.

As many of our variables of interest were correlated with one another, we calculated partial correlations between each variable and expression divergence while controlling for the other variables. The set of variables included are described in Supplementary Information, Section 1.3. We only used filtered genes for which we had information on all the variables ($n = 1,832$). Partial correlations were calculated from Spearman's rank correlation matrices using the R package 'corpcor'. Ninety-five per cent confidence intervals for each partial correlation were generated by boot-strapping (random sample with replacement) the set of genes contributing to the correlation. One thousand bootstraps were performed using the R package 'boot'.

31. Markow, T. A. & O'Grady, P. M. *Drosophila* biology in the genomic age. *Genetics* **177,** 1269–1276 (2007).
32. Ahmed, N., Natarajan, T. & Rao, K. R. Discrete cosine transform. *IEEE Trans. Comput.* **C-23,** 90–93 (1974).
33. Kerr, M. K., Martin, M. & Churchill, G. A. Analysis of variance for gene expression microarray data. *J. Comput. Biol.* **7,** 819–837 (2000).
34. Smyth, G. K. Limma: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions using R and Bioconductor* 397–420 (Springer, 2005).
35. Smyth, G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3,** Article3 (2004).
36. Irizarry, R. A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31,** e15 (2003).
37. Jonnalagadda, S. & Srinivasan, R. Principal components analysis based methodology to identify differentially expressed genes in time-course microarray data. *BMC Bioinformatics* **9,** 267 (2008).
38. Felsenstein, J. PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics* **5,** 164–166 (1989).
39. Huson, D. H. *et al.* Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics* **8,** 460 (2007).
40. Heger, A. & Ponting, C. P. Evolutionary rate analyses of orthologs and paralogs from 12 *Drosophila* genomes. *Genome Res.* **17,** 1837–1849 (2007).
41. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20,** 289–290 (2004).
42. Hansen, T. F. & Martins, E. P. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* **50,** 1404–1417 (1996).
43. Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22,** 1540–1542 (2006).
44. Ge, Y., Dudoit, S. & Speed, T. P. *Resampling-Based Multiple Testing for Microarray Data Analysis.* Technical Report (Univ. California, 2003).
45. Alexa, A., Rahnenfhrer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22,** 1600–1607 (2006).
46. Grossmann, S., Bauer, S., Robinson, P. N. & Vingron, M. Improved detection of overrepresentation of Gene-Ontology annotations with parent child analysis. *Bioinformatics* **23,** 3024–3031 (2007).
47. Agudelo-Romero, P. *et al.* Changes in the gene expression profile of *Arabidopsis thaliana* after infection with Tobacco etch virus. *Virol. J.* **5,** 92 (2008).
48. Larracuente, A. M. *et al.* Evolution of protein-coding genes in *Drosophila. Trends Genet.* **24,** 114–123 (2008).
49. Smith, C. D., Shu, S., Mungall, C. J. & Karpen, G. H. The Release 5.1 annotation of *Drosophila melanogaster* heterochromatin. *Science* **316,** 1586–1591 (2007).
50. Chintapalli, V. R., Wang, J. & Dow, J. A. T. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nature Genet.* **39,** 715–720 (2007).

doi:10.1038/nature09632

A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns

Tomislav Domazet-Lošo[1,2] & Diethard Tautz[1]

Parallels between phylogeny and ontogeny have been discussed for almost two centuries, and a number of theories have been proposed to explain such patterns[1]. Especially elusive is the phylotypic stage, a phase during development where species within a phylum are particularly similar to each other[2–6]. Although this has formerly been interpreted as a recapitulation of phylogeny[1], it is now thought to reflect an ontogenetic progression phase[2], where strong constraints on developmental regulation and gene interactions exist[2,3]. Several studies have shown that genes expressed during this stage evolve at a slower rate, but it has so far not been possible to derive an unequivocal molecular signature associated with this stage[7–15]. Here we use a combination of phylostratigraphy[16] and stage-specific gene expression data to generate a cumulative index that reflects the evolutionary age of the transcriptome at given ontogenetic stages. Using zebrafish ontogeny and adult development as a model, we find that the phylotypic stage does indeed express the oldest transcriptome set and that younger sets are expressed during early and late development, thus faithfully mirroring the hourglass model of morphological divergence[2,3]. Reproductively active animals show the youngest transcriptome, with major differences between males and females. Notably, ageing animals express increasingly older genes. Comparisons with similar data sets from flies and nematodes show that this pattern occurs across phyla. Our results indicate that an old transcriptome marks the phylotypic phase and that phylogenetic differences at other ontogenetic stages correlate with the expression of newly evolved genes.

The evolutionary origin of genes can be traced by similarity searches in genomes representing the whole tree of life. We have called this approach 'phylostratigraphy' and have shown that meaningful comparisons can be derived from it[16–18] (see Supplementary Note 1). It is important to note that the procedure identifies specifically the origin of novel genes with no traceable relation to existing genes or protein domains (see Supplementary Note 2). Another important property of phylostratigraphy is that it establishes a phylogenetic scale where every gene within a genome has its phylogenetic rank. Here, using this phylogenetic hierarchy, we extend this approach by linking it to all expressed genes within the ontogenetic sequence. To link these two hierarchies quantitatively we developed a transcriptome age index (TAI), which integrates the age of a gene with its expression level at a given developmental stage and sums this over all genes expressed at the respective stage. The higher the TAI, the younger the transcriptome (see Methods).

To apply the TAI for a developmental model system, we have generated a fine-grained series of transcriptome data of zebrafish development, covering a total of 60 stages, from unfertilized eggs to ageing animals. Figure 1a shows the TAI profile, plotted along these stages. The comparatively oldest transcript sets are expressed during the late segmentation/early pharyngula stage, which is the developmental stage that is usually equated with the phylotypic stage in zebrafish[13]. The start of heart pulsations and blood circulation in the embryo (24 h)[19] is a morphological feature that approximately marks this period of lowest TAI values. Phylogenetically younger transcriptome sets are expressed before and after this stage. This correlates well with the observation that early and late stages of chordate development also show a higher morphological divergence between taxa[2,3]. During the mid-larval stage we see a second phase where older transcriptomes are expressed, which corresponds to metamorphosis[19]. Although metamorphosis in fish is not as overt as in some other chordates (for example, amphibians), it is nonetheless a phase with major changes in morphology and life-history strategy. It is particularly evident in the reshaping of the fins, which change from a basal pattern that is seen across all fish into the one that is more specific for zebrafish[20]. After this stage, the transcriptome becomes younger and peaks in young adults. Males and females show major differences in the overall age index, with females expressing the relatively youngest genes. Intriguingly, as animals become older, they express older genes again.

Analysing the contribution of the different phylostrata (ps) to the general profile shows that they contribute to different extents (Fig. 1b). Genes that have emerged before the evolution of metazoa (ps1 to ps5) are more equally expressed throughout ontogeny, whereas later-emerging genes contribute increasingly to the differential pattern. A more detailed contribution of the genes from the different phylostrata is summarized in Fig. 2. Here we have depicted the relative expression levels for each stage for several phylostrata. This representation is only partly comparable to that in Fig. 1b, as it disregards the actual number of genes within a phylostratum. But this analysis allows several more specific points to be made.

Most genes that have arisen in ps1 (cellular origin) are general enzymes and housekeeping genes, but their RNA is not highly expressed before gastrulation (Fig. 2a), indicating that the products of these genes are primarily stored as proteins in the egg. Intriguingly, this is very different for the genes from ps2 to ps4, which have their relatively highest expression levels at these early stages. This is also indicative of a correlation between phylogenetic age of a gene and ontogenetic use of its product.

The noticeable TAI peak during gastrulation (Fig. 1a) is mainly generated by the genes from ps5 (evolution of metazoa, Fig. 2b). Studies in sponges suggest that gastrulation is an embryological process present since the onset of the metazoan evolution[21], which is in agreement with the peak of ps5 genes. In addition, we have previously identified ps5 as the time of emergence of genes involved in cellular interactions[18], which are evidently of particular importance during gastrulation.

Genes that have evolved during chordate evolution (ps9) are particularly highly expressed at the end of the pharyngula stage and at the beginning of larval stages, before metamorphosis (Fig. 2b). This is again a very suggestive correlation, because during this phase the chordate body plan in zebrafish reaches, for the first time, a full functional differentiation that is reflected in chordate-specific undulatory swimming and the start of active feeding. Interestingly, ps7 genes (evolution of bilateria) start to be strongly expressed at the beginning

[1]Max-Planck-Institut für Evolutionsbiologie, August-Thienemannstrasse 2, 24306 Plön, Germany. [2]Laboratory of Evolutionary Genetics, Division of Molecular Biology, Ruđer Bošković Institute, Bijenička cesta 54, P.P. 180, 10002 Zagreb, Croatia.

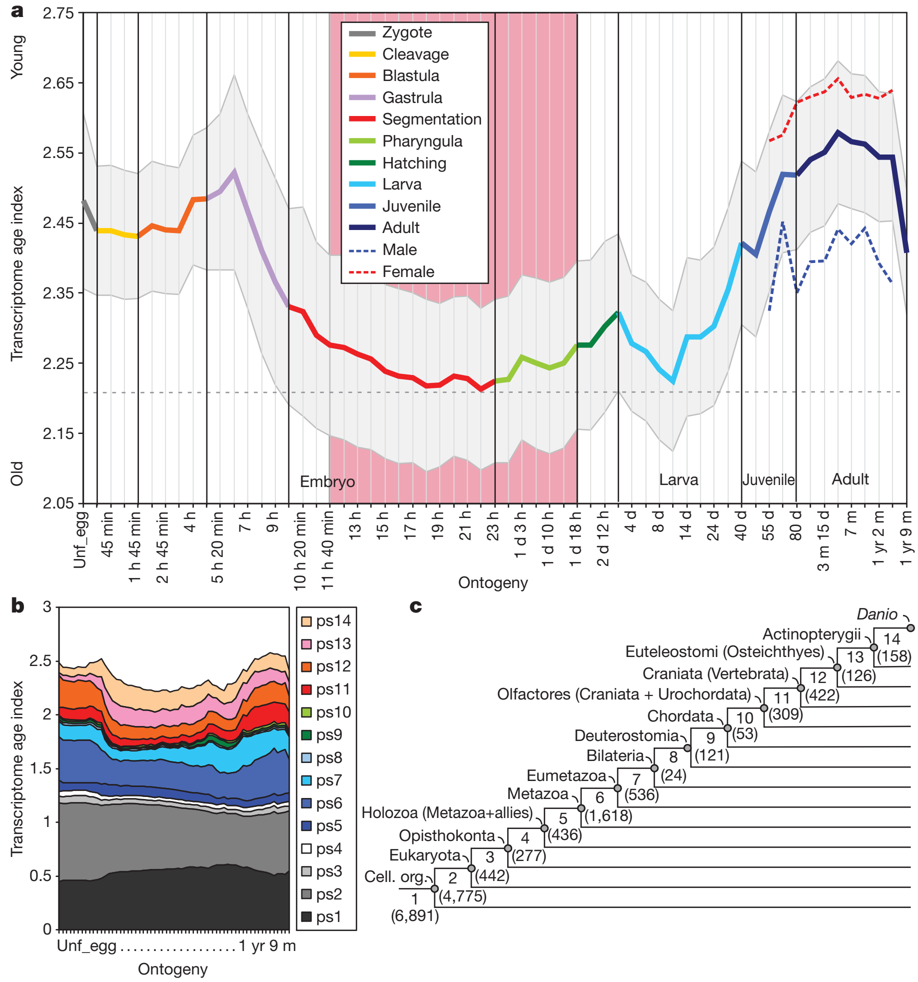


Figure 1 | Transcriptome age profiles for the zebrafish ontogeny. **a**, Cumulative transcriptome age index (TAI) for the different developmental stages. The pink shaded area represents the presumptive phylotypic phase in vertebrates. The overall pattern is significant by repeated measures ANOVA ($P = 2.4 \times 10^{-15}$, after Greenhouse–Geisser correction $P = 0.024$). Grey shaded areas represent ± the standard error of TAI estimated by bootstrap analysis. **b**, Transcriptome indices split according to the origin of the genes from the different phylostrata, based on the same developmental series as in **a**. **c**, Depiction of the phylostrata analysed; numbers in parentheses denote the number of array probes analysed for each phylostratum.

of metamorphosis, raising the possibility that metamorphosis-related genes or processes have already originated in parallel to the formation of bilateria. Ancient origins of hormonal signalling processes associated with metamorphosis have indeed been proposed[22].

Although comparable fine-grained data sets are currently not available for other model systems, one can still compare the trends based on available partial data sets. A good developmental transcriptome data series exists for *Drosophila*, although it covers only one-third of the expressed genes[23]. We have calculated the TAI for these data and find that the overall pattern is indeed comparable to zebrafish (Fig. 3). Most notably, the relatively oldest transcriptome is expressed during germ-band elongation, which can be equated to the phylotypic phase in arthropods[24]. Thus, this molecular signature is qualitatively comparable to the zebrafish data, but there are more novel genes among the post-embryonically expressed genes in *Drosophila* than in zebrafish, reflected in larger TAI values from differentiation stages onwards (Fig. 3a). Again, we see a major difference between males and females

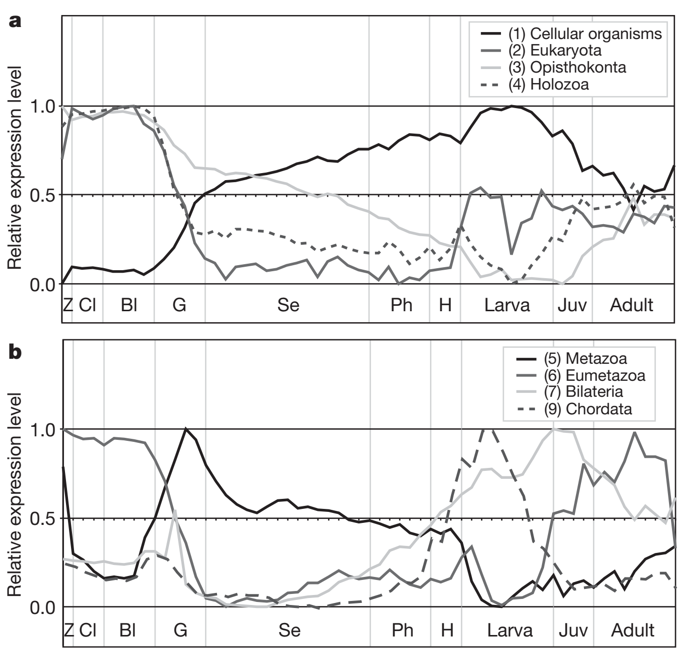


Figure 2 | Relative expression of the genes from each phylostratum across the zebrafish ontogeny (same stages as in Fig. 1) for selected phylostrata with significant differences. See Supplementary Fig. 3 for representation of all phylostrata and significance assessments. For easier comparisons, the relative expression calculated in relation to the highest (0) and lowest (1) expression values across developmental stages is shown (see Methods). Bl, blastula; Cl, cleavage; G, gastrula; H, hatching; Juv, juvenile; Ph, pharyngula; Se, segmentation; Z, zygote.

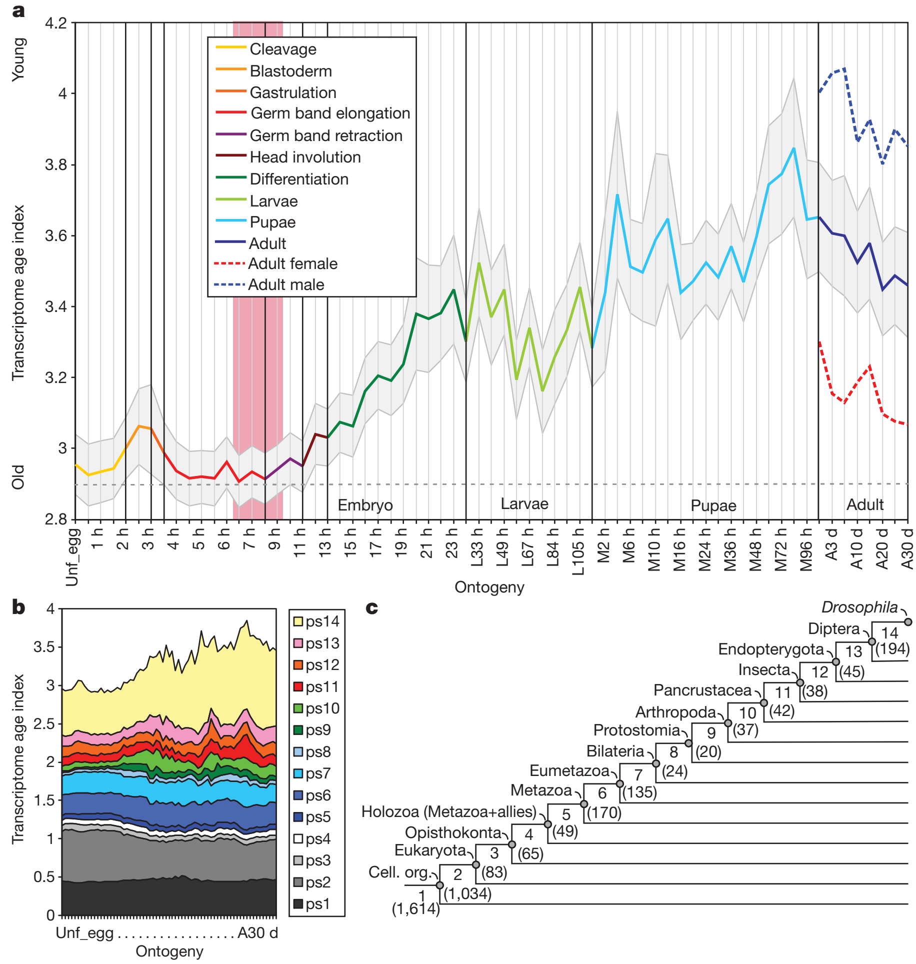


Figure 3 | Transcriptome age profiles for the *Drosophila* ontogeny, based on the data in ref. 23. **a**, Cumulative transcriptome index for the different developmental stages. The pink shaded area represents the presumptive phylotypic phase in insects. The overall pattern of differences in TAI is significant by repeated measures ANOVA ($P = 2.5 \times 10^{-93}$, after Greenhouse–Geisser correction $P = 1.22 \times 10^{-11}$). Grey shaded areas represent ± the standard error of TAI estimated by bootstrap analysis. **b**, **c** Same as for Fig. 1b, c.

after hatching, but in contrast to zebrafish, the males express the younger transcriptomes. Similar to the situation in zebrafish, ageing *Drosophila* express increasingly older transcriptomes (Fig. 3a). Breaking this pattern down to the contribution from the different phylostrata shows again that the oldest genes contribute little to the differential pattern, whereas genes that have emerged in ps9 (equivalent to the evolution of Arthropods) and later add increasingly to the final profile (Fig. 3b).

Comparable, but even more limited, ontogenetic transcriptome data are also available for the nematode *Caenorhabditis elegans*[25] and the mosquito *Anopheles*[26]. The same trends can be seen for those as well, namely the oldest genes expressed during the embryonic stages, the youngest towards adult stages and older genes in ageing animals (Supplementary Figs 1 and 2).

These consistent overall patterns across phyla, as well as the detailed analysis within zebrafish, suggest that there is a link between evolutionary innovations and the emergence of novel genes[16,27,28]. Adaptations are expected to occur primarily in response to altered ecological conditions. Juvenile and adults interact much more with ecological factors than embryos, which may even be a cause for fast postzygotic isolation[29]. Similarly, the zygote may also react to environmental constraints, for example, via the amount of yolk provided in the egg. In contrast, mid-embryonic stages around the phylotypic phase are normally not in direct contact with the environment and are therefore less likely to be subject to ecological adaptations and evolutionary change. As already suggested by Darwin (discussed in ref. 15), this alone could explain the lowered morphological divergence of early ontogenetic stages compared to adults, which would obviate the need to invoke particular constraints. Alternatively, the constraint hypothesis would suggest that it is difficult for newly evolved genes to become recruited to strongly connected regulatory networks[12,13,15].

The fact that ageing animals revert to older transcriptomes is in line with the notion that animals beyond the reproductive age are not 'visible' to natural selection and can therefore not be subject to specific adaptations any more. Also, the fact that major TAI differences can be seen between males and females could have been anticipated, because sexual selection is expected to continuously change phenotypic traits between them. However, the fact that the differences go in opposite directions in zebrafish and *Drosophila* is surprising. We have therefore studied in detail which phylostrata contribute most to these differences (Fig. 4). Both taxa show a female expression bias of ps2 genes, which may be correlated to egg production, as RNA from such genes is stored in the eggs (see above). But they strongly deviate at other phylostrata. Zebrafish shows a strong female bias of ps6 and ps12 genes, which is absent in *Drosophila* (Fig. 4). *Drosophila*, on the other hand, shows an extreme bias of ps14 genes in males (Fig. 4), which is caused by the many orphan genes involved in spermatogenesis[30]. Thus, in contrast to the ontogenetic similarities of the TAI trends between the two taxa, the sex differences are rather incongruent and indicate different evolutionary trajectories for male–female differences.

Our study provides strong molecular support for a correlate between phylogeny and ontogeny, as well as the hourglass model of

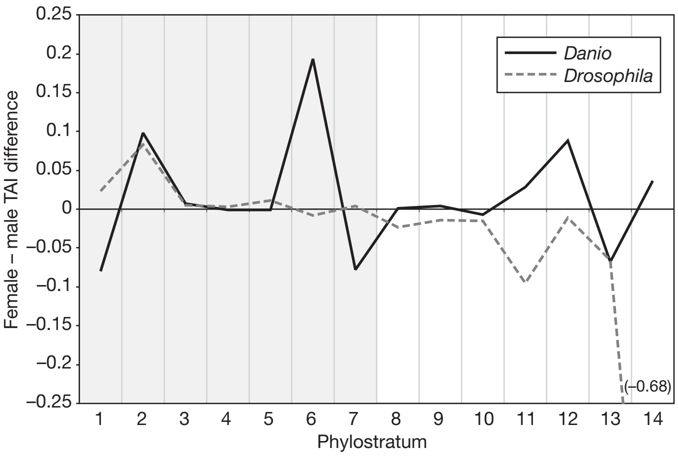


Figure 4 | Comparison of differences in TAI between females and males. Comparison across phylostrata in zebrafish (*Danio*) and *Drosophila* (see Supplementary Fig. 4 for a plot that includes the differences between stages). The grey shaded area designates the shared part of the phylogeny between the two species (origin of the first cell until the last common ancestor of Bilateria, ps1–ps7). Note that the ps14 value for *Drosophila* is off the scale (difference is given in parenthesis).

development. Under this scheme, the phylotypic phase can be defined as the ontogenetic progression during which the oldest gene set is expressed, either because this is the phase with the lowest opportunity for lineage-specific adaptations, or because it is internally so constrained that newly evolved genes cannot become integrated.

METHODS SUMMARY

The TAI is the weighted mean of phylogenetic ranks (phylostrata) and is calculated for every ontogenetic stage *s* as follows:

$$\mathrm{TAI}_S = \frac{\sum_{i=1}^{n} \mathrm{ps}_i e_i}{\sum_{i=1}^{n} e_i}$$

where ps_i is an integer that represents the phylostratum of the gene *i* (for example, 1, the oldest; 14, the youngest), e_i is the microarray signal intensity value (obtained from Agilent Zebrafish (V2) Gene Expression Microarrays) of the gene *i* that acts as weight factor and *n* is the total number of genes analysed. This way of calculating the index gives an increasingly stronger weight to younger phylostrata, thus compensating for the fact that the older phylostrata usually harbour the larger number of genes[16–18].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 September; accepted 27 October 2010.

1. Gould, S. J. *Ontogeny and Phylogeny* (Harvard Univ. Press, 1977).
2. Duboule, D. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Dev. Suppl.* 135–142 (1994).
3. Raff, R. A. *The Shape of Life: Genes Development, and the Evolution of Animal Form* (Univ. Chicago Press, 1996).
4. Richardson, M. K. *et al.* There is no highly conserved embryonic stage in the vertebrates: implications for current theories of evolution and development. *Anat. Embryol. (Berl.)* **196,** 91–106 (1997).
5. Hall, B. K. Phylotypic stage or phantom: is there a highly conserved embryonic stage in vertebrates? *Trends Ecol. Evol.* **12,** 461–463 (1997).
6. Bininda-Emonds, O. R., Jeffery, J. E. & Richardson, M. K. Inverting the hourglass: quantitative evidence against the phylotypic stage in vertebrate development. *Proc. R. Soc. Lond. B* **270,** 341–346 (2003).
7. Hanada, K., Shiu, S. H. & Li, W. H. The Nonsynonymous/synonymous substitution rate ratio versus the radical/conservative replacement rate ratio in the evolution of mammalian genes. *Mol. Biol. Evol.* **24,** 2235–2241 (2007).
8. Comte, A., Roux, J. & Robinson-Rechavi, M. Molecular signaling in zebrafish development and the vertebrate phylotypic period. *Evol. Dev.* **12,** 144–156 (2010).
9. Yassin, A., Lienau, E. K., Narechania, A. & DeSalle, R. Catching the phylogenic history through the ontogenic hourglass: a phylogenomic analysis of *Drosophila* body segmentation genes. *Evol. Dev.* **12,** 288–295 (2010).
10. Davis, J. C., Brandman, O. & Petrov, D. A. Protein evolution in the context of *Drosophila* development. *J. Mol. Evol.* **60,** 774–785 (2005).
11. Hazkani-Covo, E., Wool, D. & Graur, D. In search of the vertebrate phylotypic stage: a molecular examination of the developmental hourglass model and von Baer's third law. *J. Exp. Zoolog. B* **304,** 150–158 (2005).
12. Irie, N. & Sehara-Fujisawa, A. The vertebrate phylotypic stage and an early bilaterian-related stage in mouse embryogenesis defined by genomic information. *BMC Biol.* **5,** 1 (2007).
13. Roux, J. & Robinson-Rechavi, M. Developmental constraints on vertebrate genome evolution. *PLoS Genet.* **4,** e1000311 (2008).
14. Cruickshank, T. & Wade, M. J. Microevolutionary support for a developmental hourglass: gene expression patterns shape sequence variation and divergence in *Drosophila*. *Evol. Dev.* **10,** 583–590 (2008).
15. Artieri, C. G., Haerty, W. & Singh, R. S. Ontogeny and phylogeny: molecular signatures of selection, constraint, and temporal pleiotropy in the development of *Drosophila*. *BMC Biol.* **7,** 42 (2009).
16. Domazet-Lošo, T., Brajković, J. & Tautz, D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* **23,** 533–539 (2007).
17. Domazet-Lošo, T. & Tautz, D. An ancient evolutionary origin of genes associated with human genetic diseases. *Mol. Biol. Evol.* **25,** 2699–2707 (2008).
18. Domazet-Lošo, T. & Tautz, D. Phylostratigraphic tracking of cancer genes suggests a link to the emergence of multicellularity in metazoa. *BMC Biol.* **8,** 66 (2010).
19. Parichy, D. M., Elizondo, M. R., Mills, M. G., Gordon, T. N. & Engeszer, R. E. Normal table of postembryonic zebrafish development: staging by externally visible anatomy of the living fish. *Dev. Dyn.* **238,** 2975–3015 (2009).
20. Grandel, H. & Schulte-Merker, S. The development of the paired fins in the zebrafish (*Danio rerio*). *Mech. Dev.* **79,** 99–120 (1998).
21. Leys, S. P. & Eerkes-Medrano, D. Gastrulation in calcareous sponges: In search of Haeckel's Gastraea. *Integr. Comp. Biol.* **45,** 342–351 (2005).
22. Flatt, T., Moroz, L. L., Tatar, M. & Heyland, A. Comparing thyroid and insect hormone signaling. *Integr. Comp. Biol.* **46,** 777–794 (2006).
23. Arbeitman, M. N. *et al.* Gene expression during the life cycle of *Drosophila melanogaster*. *Science* **297,** 2270–2275 (2002).
24. Sander, K. In *Development and Evolution: the sixth Symposium of the British Society for Developmental Biology* (eds Goodwin, B. C., Holder, N. & Wylie, C. C.) 137–159 (Cambridge Univ. Press, 1983).
25. Hill, A. A., Hunter, C. P., Tsung, B. T., Tucker-Kellogg, G. & Brown, E. L. Genomic analysis of gene expression in *C. elegans*. *Science* **290,** 809–812 (2000).
26. Koutsos, A. C. *et al.* Life cycle transcriptome of the malaria mosquito *Anopheles gambiae* and comparison with the fruitfly *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **104,** 11304–11309 (2007).
27. Domazet-Lošo, T. & Tautz, D. An evolutionary analysis of orphan genes in *Drosophila*. *Genome Res.* **13,** 2213–2219 (2003).
28. Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R. & Bosch, T. C. More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet.* **25,** 404–413 (2009).
29. Nolte, A. W., Renaut, S. & Bernatchez, L. Divergence in gene regulation at young life history stages of whitefish (*Coregonus* sp.) and the emergence of genomic isolation. *BMC Evol. Biol.* **9,** 59 (2009).
30. Levine, M. T., Jones, C. D., Kern, A. D., Lindfors, H. A. & Begun, D. J. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and show testis-biased expression. *Proc. Natl Acad. Sci. USA* **103,** 9935–9939 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Walderich for providing zebrafish, A. Nolte, E. Blohm-Sievers, B. Kleinhenz, L. Turner and J. Bryk for laboratory support, R. Bakarić has provided the phylostratigraphic map of *C. elegans*, and M. Domazet-Lošo and V. Dunjko have helped with statistics. L. Boell, F. Chang and A. Pozhitkov have made suggestions on the manuscript. This work was supported by Unity Through Knowledge Fund (grant No. 49), Adris Foundation and funds of the Max-Planck Society. Computational resources were provided by CSTMB and RBI (Phylostrat Cluster).

Author Contributions T.D.-L. conceived the basic idea and conducted the experiments; D.T. contributed to the evaluation and interpretation of the results. Both authors wrote the manuscript.

Author Information The microarray data for zebrafish were deposited at the NCBI Gene Expression Omnibus (GEO) repository under the accession number GSE24616. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.D.-L. (tdomazet@irb.hr).

METHODS

Fish keeping and sampling conditions. Zebrafish (*Danio rerio*) were kept in 12 l flow-through tanks at 26.5 °C (around 60 animals per tank). For accurate staging, fertilized eggs were collected within 15-min intervals and incubated in Petri dishes at 28.5 °C with water changes every 2–6 h. After hatching, larvae were transferred to 1-l tanks and kept at 28.5 °C. We took, in total, 72 samples in two replicates that correspond to 60 stages across zebrafish ontogeny (50 samples before the sex could be clearly recognized plus 11 samples of males and females each). Staging was done according to post-fertilization time. Embryos were additionally staged under the dissecting microscope according to ref. 31 to check for the consistency of post-fertilization timing and morphological development at standard temperature (28.5 °C). Only healthy animals that showed the expected morphological features for a given post-fertilization time were sampled. Each sample contained around 50 individuals until the 1 day and 3 h embryo stage, 15 individuals until the 10 day larval stage, 10 individuals until the 18 day larval stage, 5 individuals until the 45 day juvenile stage, whereas in later juvenile and adult stages we sampled males and females separately and each sample contained 2 individuals. All samples were snap frozen in liquid nitrogen and stored at −80 °C until RNA extraction. To avoid severe biases owing to the excess of unfertilized eggs, we squeezed eggs from adult females before freezing them in liquid nitrogen.

Phylostratigraphy. A full account of phylostratigraphic analysis and theoretical underpinnings has been presented previously[16–18]. The zebrafish genes of the present study (28,546, ENSEMBL release 56) were mapped on the currently best supported phylogeny using BLAST searches against the cleaned up and additionally enriched NCBI NR database, which represents the most exhaustive set of known proteins across all organisms. Our choice of internodes (phylostrata) in the consensus phylogeny depended on the availability of complete annotated genomes, reliability of phylogenetic relationships and on the importance of evolutionary transitions. Similarly, the data of *Drosophila* (13,389 genes), *Anopheles* (12,457 genes) and *Caenorhabditis elegans* (19,077 genes) were mapped to the best supported phylogenies that represent their evolutionary lineages.

RNA isolation and microarray gene expression experiments. Total RNA was isolated using the TRIZOL plus protocol (Invitrogen). Four-hundred nanograms of total RNA per sample were Cy3 labelled according to the one-colour Quick Amp Labelling Kit protocol (Agilent). Labelled cRNAs were hybridized to Agilent Zebrafish (V2) Gene Expression Microarray slides (4 × 44k) for 17 h at 65 °C and washed according to the Agilent protocol. Hybridized microarray slides were scanned using an Agilent High-Resolution Microarray Scanner.

Microarray data extraction, filtering and analysis. Raw microarray image files were processed and quality checked by Agilent's Feature Extraction 10.7 Image Analysis Software. Background subtracted signal intensity values that contain correction for multiplicative surface trends (gProcessedSignal) generated by Feature Extraction Software were used for further data analysis. Using GeneSpring microarray data analysis software we filtered probes that were flagged as non-uniform or as population outlier. For every of the 72 samples we calculated average signal intensity values over the two biological replicates. Probes (60 bp) were mapped on the *Danio rerio* transcripts (ENSEMBL version 54) that passed the phylostratigraphic analysis (see below) using CD-hit software. This procedure yielded 16,188 unique probes that collapsed to 12,892 ENSEMBL predicted genes.

Phylostratigraphically mapped genes of *Drosophila* were linked to available microarray data[23]. This procedure yielded a data set of 3,550 genes. In a similar fashion, phylogenetically ranked microarray data sets were obtained for *C. elegans*[25] (16,832 genes) and *Anopheles*[26] (3,135 genes).

Transcriptome age index and statistical analysis. The TAI is the weighted mean of phylogenetic ranks (phylostrata) and is calculated for every ontogenetic stage *s* as follows:

$$\mathrm{TAI}_S = \frac{\sum_{i=1}^{n} \mathrm{ps}_i e_i}{\sum_{i=1}^{n} e_i}$$

where ps_i is an integer that represents the phylostratum of the gene *i* (for example, 1, the oldest; 14, the youngest), e_i is the microarray signal intensity value (obtained from Agilent Zebrafish (V2) Gene Expression Microarrays) of the gene *i* that acts as weight factor and *n* is the total number of genes analysed. This way of calculating the index gives an increasingly stronger weight to younger phylostrata, thus compensating for the fact that the older phylostrata usually harbour the larger number of genes[16–18].

We chose to calculate the TAI index based on the amount of expression per gene, rather than by simply adding up whether a gene is expressed or not. Although this latter approach would also seem feasible, it runs into a technical problem. To say that a given gene is expressed or not, one would have to impose a cutoff on the signals from the microarrays, which is more or less arbitrary, as a weak signal on a microarray could be derived from a gene with very low expression level, or from a highly expressed gene that is present in a few cells only. Also, absolute quantities are difficult to compare across microarrays and a single cutoff value would not be appropriate (see below). In balance, we have therefore opted for the expression level as a numerator, also because one could argue that genes that are broadly expressed at high levels should be more relevant than specialized genes.

The TAI formula can alternatively be written as:

$$\mathrm{TAI}_S = \mathrm{ps}_1 \frac{e_1}{e_1+e_2+\ldots+e_n} + \mathrm{ps}_2 \frac{e_2}{e_1+e_2+\ldots+e_n} + \ldots + \mathrm{ps}_n \frac{e_n}{e_1+e_2+\ldots+e_n}$$

The expression $e_1 + e_2 + \ldots + e_n$ represents the total signal of the analysed probes on the microarray, whereas the ratio $e_i/(e_1 + e_2 + \ldots + e_n)$, which can be denoted as f_i, represents the partial concentration (frequency) of probe *i* in the total microarray signal at a given stage; it is within a range between zero and one. It is important to note that the calculation of the partial concentration (f_i) inherently makes a global intensity normalization over the microarray experiment at a given stage and that at every stage the sum of partial concentrations will equal one. In many microarray studies it is common to assess the direction of expression change (over- or under-expression). This type of analysis requires that after the normalization procedure, which aims to remove noise from the experiment, expression signals that are measured across experiments still reflect absolute number of mRNA molecules per unit of biological material. In such situations, if global intensity normalization is applied, it must be assumed that the total number of mRNA copies for all genes on the array does not significantly differ between experiments. Contrary to this common application of microarrays, in our study we are not interested in the direction of expression change of particular genes. Instead, we are looking at how partial concentrations of RNAs contribute to the overall transcriptome across stages. For this purpose it is irrelevant which part of the transcriptome is responsible for change of the partial concentration. Therefore in our data treatment it is not necessary to assume that cumulative signals do not differ between experiments. This shift in perspective greatly simplifies the analysis on the scale of the complete ontogeny because abundance and distribution of transcripts is commonly very different between stages.

Thus, the TAI can be written as a sum of products between partial concentration and corresponding phylostratum:

$$\mathrm{TAI}_S = \sum_{i=1}^{n} \mathrm{ps}_i f_i = \mathrm{ps}_1 f_1 + \mathrm{ps}_2 f_2 + \ldots + \mathrm{ps}_n f_n$$

To asses the contributions of a specific phylostratum to the overall TAI (Figs 1b and 3b) we split the above total sum of $\mathrm{ps}_i f_i$ products to subsets of $\mathrm{ps}_i f_i$ sums where the value of ps_i (phylostratigraphic rank) was used as a grouping factor.

By applying repeated measures ANOVA on these $\mathrm{ps}_i f_i$ products we tested the significance of difference in TAI between stages. Repeated measures ANOVA was used because the same set of probes are measured at every stage, that is, there is dependence between the stages compared. Before means of these products across stages are compared by ANOVA we multiplied every $ps_i f_i$ product with constant *n* (total number of analysed probes). This transformation does not influence the ANOVA analysis and its sole purpose is that means of $\mathrm{ps}_i f_i$ products compared in ANOVA are equal to the corresponding TAI values. Because the assumption of sphericity was violated in the data sets analysed by repeated measures ANOVA, we applied the Greenhouse–Geisser correction. Multivariate test statistics, an alternative approach that is not dependent on the assumption of sphericity, corroborated our statistical results of repeated measures ANOVA. We used the bootstrap approach (1,000 replicates) to asses the standard error of weighted mean (TAI)[32].

Relative expression of the genes for a given phylostratum (ps) and developmental stage (*s*) (Fig. 2) was calculated according to the equation:

$$\mathrm{RE(ps)}_s = \frac{\bar{f} - \bar{f}_{\min}}{\bar{f}_{\max} - \bar{f}_{\min}}$$

where $\bar{f}$ is the average partial concentration of RNAs from phylostratum ps for a given stage and $\bar{f}_{\max}, \bar{f}_{\min}$ are the maximal and minimal average partial concentration from phylostratum ps across all considered stages, respectively.

31. Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages of embryonic development of the zebrafish. *Dev. Dyn.* **203,** 253–310 (1995).
32. Efron, B. & Tibshirani, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* **1,** 54–75 (1986).

doi:10.1038/nature09551

Interdependence of behavioural variability and response to small stimuli in bacteria

Heungwon Park[1], William Pontius[2], Calin C. Guet[3], John F. Marko[4], Thierry Emonet[2] & Philippe Cluzel[3]

The chemotaxis signalling network in *Escherichia coli* that controls the locomotion of bacteria is a classic model system for signal transduction[1,2]. This pathway modulates the behaviour of flagellar motors to propel bacteria towards sources of chemical attractants. Although this system relaxes to a steady state in response to environmental changes, the signalling events within the chemotaxis network are noisy and cause large temporal variations of the motor behaviour even in the absence of stimulus[3]. That the same signalling network governs both behavioural variability and cellular response raises the question of whether these two traits are independent. Here, we experimentally establish a fluctuation–response relationship in the chemotaxis system of living bacteria. Using this relationship, we demonstrate the possibility of inferring the cellular response from the behavioural variability measured before stimulus. In monitoring the pre- and post-stimulus switching behaviour of individual bacterial motors, we found that variability scales linearly with the response time for different functioning states of the cell. This study highlights that the fundamental relationship between fluctuation and response is not constrained to physical systems at thermodynamic equilibrium[4] but is extensible to living cells[5]. Such a relationship not only implies that behavioural variability and cellular response can be coupled traits, but it also provides a general framework within which we can examine how the selection of a network design shapes this interdependence.

It is standard procedure to characterize the stochastic dynamics of physical systems in thermodynamic equilibrium by measuring spontaneous fluctuations and responses to small external perturbations. Because these two distinct measurements contain the same information, they are related by the fluctuation-dissipation theorem[4]. Although the fluctuation-dissipation theorem has practical applications—to evaluate force-extension sensors for single biomolecules[6,7] and to predict static cell-to-cell variability of gene expression[8,9]—it has not been possible to apply it directly to the study of the dynamical behaviour of living cells because they are open systems with significant non-thermal dynamics. However, this theorem has recently been extended to a fluctuation-response theorem (FRT) for systems that are not in thermodynamic equilibrium but that have a well-defined steady state and Markovian dynamics[5,10–12]. For application to living cells this condition amounts to studying dynamic processes with sufficiently short 'memory' that they can relax to a well-defined steady state. Here we use the FRT as an operational framework to establish the interdependence of distinct cellular traits, such as cellular fluctuations and response to a small stimulus, without relying on the biochemical details of a specific signalling pathway. To tackle this question experimentally, we used the well-characterized chemotaxis system in *E. coli*, which governs bacterial locomotion[13].

The chemotaxis network regulates the rotation direction—clockwise (CW) or counter-clockwise (CCW)—of the flagellar motors, which control the swimming direction of the cell[1,2]. One of the hallmarks of bacterial chemotaxis is adaptation. Following a stepwise stimulus, the CW bias (the probability that the motor will rotate clockwise) decreases abruptly, before slowly adapting back to its pre-stimulus level. Even when bacteria are adapted to their environment, the CW bias of individual cells fluctuates around the mean. These temporal fluctuations in CW bias reflect slow fluctuations in signalling events throughout the transduction network[14]. To verify that the bacterial chemotaxis system satisfies the FRT, we monitored both the temporal fluctuations of the CW bias before stimulus and the cellular response to a small stimulus at the single-cell level. Both quantities were obtained from the time series of CW and CCW intervals of individual motors from bacteria immobilized on a glass coverslip[15] and submerged in a motility medium that does not support growth. Such single-cell experiments are complicated by inherent cell-to-cell differences in relative chemotaxis protein concentration, leading to differences in switching dynamics (Fig. 1a). To compare cells with similar behaviour, we sorted wild-type cells according to their steady-state CW bias (Fig. 1a). These CW bias bins define different classes of cells, which, despite being genetically identical, have different dynamics and must be analysed separately[3].

First, we quantified the response in single cells by measuring the length of successive CCW intervals immediately following the stimulus. The stimulus (10 nM of aspartate) used in this study is small and close to the limit of sensitivity of the bacterial chemotaxis system[16]. At the single-cell level, the length of the first CCW interval following the small stimulus (Supplementary Fig. 1a) was distributed around the mean CCW interval length before stimulus (Supplementary Fig. 1b). Given that CCW interval length is a stochastic variable, we averaged the CCW interval lengths after stimulus between cells and found that the mean length of the first CCW interval following stimulus was slightly longer than the mean pre-stimulus CCW interval length (Fig. 1b). Therefore, we expected the response of the system to be within the linear regime, which was necessary to apply the FRT. We also tested the response of the chemotaxis system for a stimulus 100 times larger (1 μM aspartate). Surprisingly, the second CCW interval following the stimulus returned to near pre-stimulus length for both large and small attractant concentrations (Fig. 1c). Although the cellular response to stimulus extends in some cases beyond the second interval (Supplementary Fig. 1d, e), these results qualitatively indicate that the first CCW interval contains most of the chemotactic response to both small and large stimuli.

To characterize the system quantitatively, we defined the response time of a single cell as the cumulative length of post-stimulus CCW intervals that are strictly longer than the mean CCW interval length before stimulus (Fig. 1b, c and Supplementary Fig. 1e; see Methods for definition of response time). This procedure yields a reasonable estimate of the response time under the condition of small stimulus (Supplementary Fig. 2). We found that the response time averaged over CW bias bins decreased with CW bias for both small (Fig. 2a) and large stimuli (Fig. 2a, inset). Because all cells returned to their pre-stimulus behaviour

[1]The James Franck Institute, The Institute for Biophysical Dynamics, and The Department of Physics, University of Chicago, Chicago, Illinois 60637, USA. [2]Department of Molecular, Cellular, and Developmental Biology and Department of Physics, Yale University, New Haven, Connecticut 06520, USA. [3]FAS Center for Systems Biology, Department of Molecular and Cellular Biology, and School of Engineering and Applied Sciences, Harvard University, 52 Oxford Street, Cambridge, Massachusetts 02138, USA. [4]Department of Molecular Biosciences and Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA.

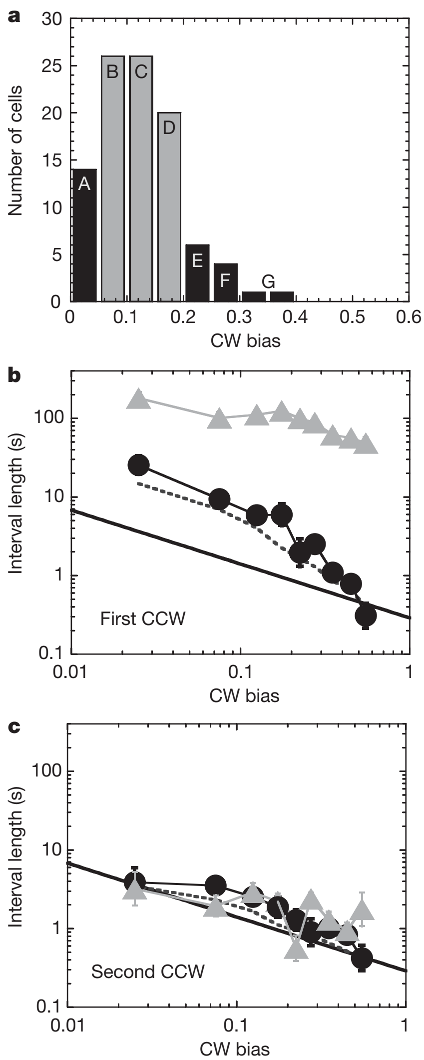


Figure 1 | CCW interval lengths pre- and post-stimulus. a, Histogram of CW bias of wild-type RP437 cells. We sorted cells into CW bias intervals by their pre-stimulus CW bias: 0.00–0.05 (A), 0.05–0.10 (B), 0.10–0.15 (C), 0.15–0.20 (D), 0.20–0.25 (E), 0.25–0.30 (F), 0.30–0.40 (G), 0.40–0.50 (H) and 0.50–0.60 (I). Grey bars are cells representative of wild-type behaviour. To increase the chance of obtaining cells with CW bias higher than 0.2, we transformed wild-type cells with pZE21-CheR (Methods). This extended the range of CW bias considered in our study to values greater than 0.4: bins H and I (not shown). **b**, **c**, The first (**b**) and second (**c**) mean post-stimulus CCW interval lengths versus pre-stimulus CW bias for all cells (wild-type RP437 and RP437 expressing CheR from pZE21-CheR). (See Supplementary Fig. 1 for individual cells.) Black circles, cells exposed to a small stimulus (10 nM L-aspartate). Grey triangles, cells exposed to a large stimulus (1 μM L-aspartate). Error bars show the standard error associated with the average CCW interval length in each bin. Dark grey dashed line, geometric mean of the CCW interval lengths following a randomly chosen time point in non-stimulated cells. Black line, power-law fit of the geometric mean of pre-stimulus CCW interval lengths calculated over 1,500 s for all cells (wild-type RP437 and RP437 expressing extra CheR from pZE21-CheR) as a function of the pre-stimulus CW bias (Supplementary Fig. 1b).

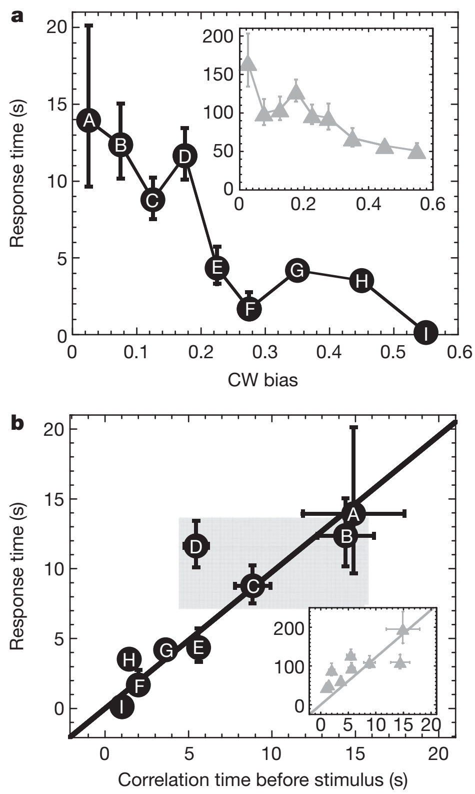


Figure 2 | Relationship between response to stimulus and fluctuations before stimulus. a, Average response time for all cells (wild-type RP437 and RP437 expressing extra CheR from pZE21-CheR) exposed to a stepwise small stimulus (black circles, 10 nM L-aspartate) or large stimulus (grey triangles in inset to **a**, 1 μM L-aspartate). The letters correspond to the CW bias bins (Fig. 1a). Error bars show the standard error associated with the average response time within each bin. **b**, Average response time to a small stimulus (black circles) or large stimulus (grey triangles in inset to **b**) as a function of the correlation time for all cells (wild-type RP437 and RP437 expressing CheR from pZE21-CheR). For the large stimulus, the average response time was adjusted by a correction factor (Supplementary Fig. 2c). The solid lines are linear fit functions forced through the origin. For the black line: response time = C × correlation time. C ≈ 0.98 ± 0.10 ($R^2 = 0.75$). For the grey line in the inset: relaxation time = C × correlation time. C ≈ 12.23 ± 1.83 ($R^2 = 0.07$). Error bars for the correlation time are the half-lengths of the first uncorrelated CCW intervals. Error bars for the response time are the standard error associated with the average response time within each bin. Grey area, representative behaviour of a wild-type population. Insets in **a** and **b** share axes with the main panels.

(Supplementary Fig. 1), the system exhibited near-precise adaptation at the single-cell level, regardless of CW bias (Supplementary Fig. 3). This result agrees with that obtained from population measurements[17,18] and shows that the dynamics have sufficiently short 'memory' and that individual cells can relax to a well-defined steady state.

A direct consequence of the linear approximation is that the response time of the system to a small external stimulus should be proportional to the correlation time of the spontaneous fluctuations before stimulus. Using serial correlation analysis[19,20], we evaluated the correlation time in non-stimulated cells (Supplementary Fig. 4). In agreement with our assumption of linear dynamics[21] and the general prediction of the FRT, we found that the correlation time scales linearly with the response time to small stimulus ($R^2 = 0.75$; Fig. 2b) whereas to large stimulus it scales poorly ($R^2 = 0.07$; Fig. 2b, inset). This result has an important practical implication: The response time that governs the cellular response in chemotaxis can be experimentally inferred by measuring the temporal correlations in behavioural fluctuations from cells before stimulus.

Cellular behavioural variability can also be defined by the amplitude of the noise rather than its temporal correlations. To characterize the amplitude of the output noise of the chemotaxis network, we computed the power spectral density of the switching binary time series measured from individual motors before stimulus (Fig. 3a and Supplementary Fig. 5). We evaluated the low-frequency noise by integrating the power spectrum between $f = 1/1{,}500\ s^{-1}$ and $f = 1/10\ s^{-1}$. In this frequency range, the temporal fluctuations are putatively caused by the slow methylation–demethylation of the receptor-kinase complexes that are also controlling the adaptive process[14]. Two elements

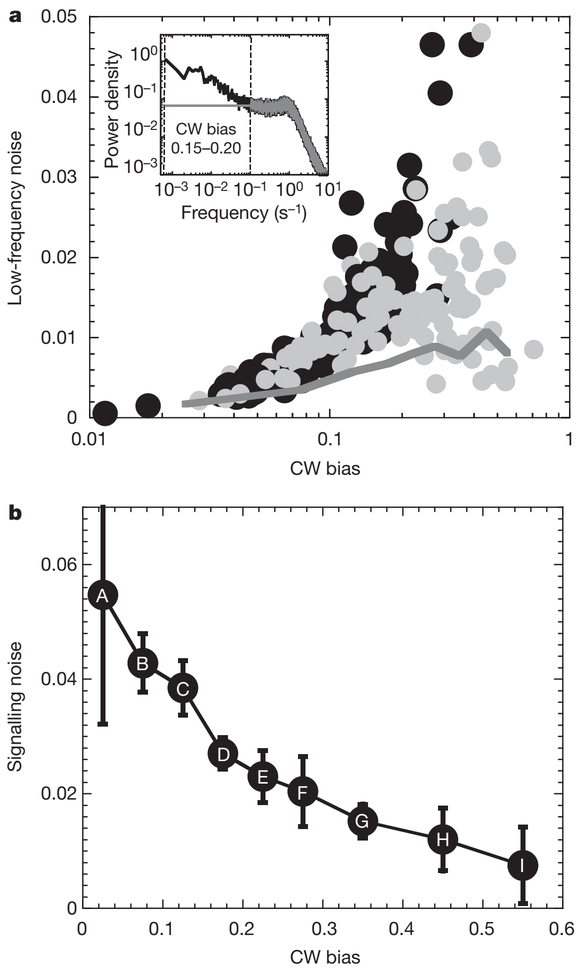


Figure 3 | Low-frequency noise in non-stimulated cells. **a**, Low-frequency noise in individual wild-type RP437 cells (black) and RP437 cells expressing CheR from pZE21-CheR (grey) versus CW bias. The inset shows power spectral density as a function of noise frequency. Black line, power density averaged over all cells (wild-type RP437 and RP437 expressing CheR from pZE21-CheR) with CW bias = 0.15–0.20. Dark grey line, power density of the motor decoupled from the signalling network[3]. We determined the low-frequency noise for the region between the dotted lines. See Supplementary Fig. 5 for all CW bias bins. **b**, Signalling noise as a function of CW bias for wild-type RP437 cells and RP437 cells expressing CheR from pZE21-CheR. Signalling noise is defined as the variance $\sigma^2_{\text{CheY-P}}$ of the fluctuating [CheY-P]. Letters correspond to the CW bias bins (Fig. 1a). The power spectral densities and CW biases are averaged over cells within the same CW bias. Error bars show the standard error associated with the estimated signalling noise within each bin.

contribute to the observed output noise: the spontaneous noise associated with the signalling events of the chemotaxis network and the stochastic switching behaviour of the bacterial motor (Fig. 3a). The binary nature of the switching behaviour of the motor dominates the variance of the noise and masks the signalling noise within the chemotaxis network the output signalling molecule of which is the phosphorylated form of the signalling protein CheY[1,2]. The active form, CheY-P, binds to the sensory basal part of the flagella rotary motor and induces CW rotation. Using a procedure developed by ref. 22, we decoupled the signalling noise, $\sigma^2_{\text{CheY-P}}$, from that of the motor. We then found that the signalling noise decreased with the CW bias (Fig. 3b).

Operationally, we used a simplified expression of the FRT, in which the response function of the chemotaxis system $\mu(t)$ and the autocorrelation function $C(t)$ of the spontaneous fluctuations of the cellular behaviour should be related by $\mu(t) = -K\frac{\mathrm{d}}{\mathrm{d}t}C(t)$. Here, the fluctuation–response coupling coefficient K may depend on the genetic background,

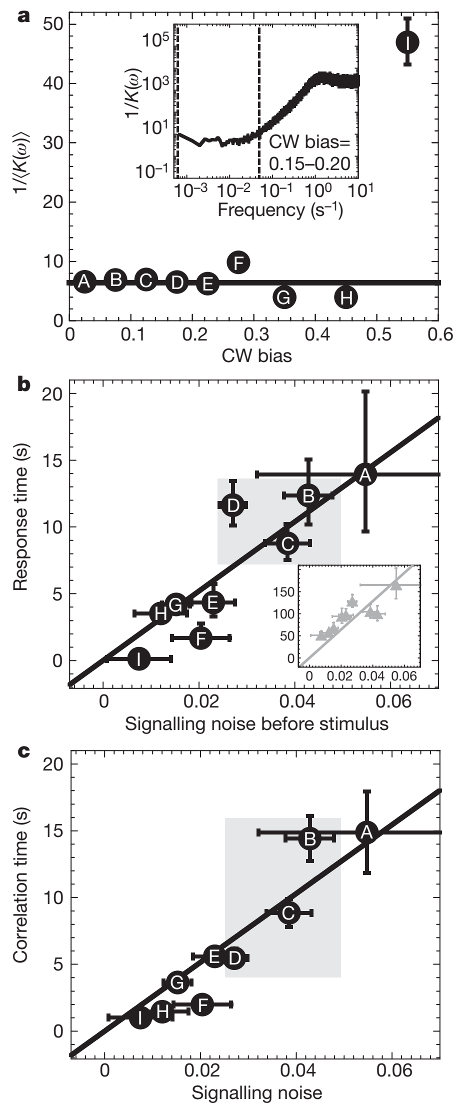


Figure 4 | Relationship between signalling noise and response time to a small external stimulus. **a**, Mean coupling coefficient $1/\langle K(\omega)\rangle_\omega$ for each CW bias bin. We computed the geometric mean over frequencies ranging from 1/1,500 s^{-1} to 1/20 s^{-1}, represented by the dashed lines in the inset to **a**. We found that the coupling coefficient K for the small stimulus was constant at long timescales for frequencies in this range (see also Supplementary Fig. 7a). The standard error of the mean is smaller than the symbol size except for the highest CW bias bin I. The line is the mean value of $1/\langle K(\omega)\rangle_\omega$ computed over CW biases ranging from 0.00 to 0.5. The inset to **a** shows $1/K(\omega)$ for cells with a CW bias ranging from 0.15 to 0.20 (10 nM L-aspartate increase). For large stimulus K is not constant (see Supplementary Fig. 7b). **b**, Average response times of all cells (wild-type RP437 and RP437 expressing inducible CheR) to small stimulus (black circles) or large stimulus (grey triangles in inset to **b**) versus mean pre-stimulus signalling noise. Solid lines are linear fits forced through the origin. Response time = $C \times \sigma^2_{\text{CheY-P}}$. Black line: $C = 259 \pm 25\,\mathrm{s}\,\mu\mathrm{M}^{-2}$ ($R^2 = 0.8$) for small stimulus. Grey line in inset to **b**: $C = 3{,}215 \pm 307\,\mathrm{s}\,\mu\mathrm{M}^{-2}$ ($R^2 = 0.4$) for large stimulus. Grey area, representative behaviour of a wild-type population. The insets in **b** shares axes with the main panel. **c**, The correlation time as a function of the mean signalling noise before stimulus for all cells (wild-type RP437 and RP437 expressing CheR from pZE21-CheR). Black line, linear fit function forced through the origin. Correlation time = $C \times \sigma^2_{\text{CheY-P}}$. $C \approx 257 \pm 21\,\mathrm{s}\,\mu\mathrm{M}^{-2}$ ($R^2 = 0.9$). Letters correspond to the CW bias bins (Fig. 1a). Error bars for the correlation time are the average half-lengths of the first uncorrelated CCW intervals. Error bars for the signalling noise are the standard error associated with the signalling noise in each bin. Grey area is representative behaviour of a wild-type population.

growth conditions, and functional state of the cell. We plotted the coefficient $K(\omega) = -\frac{2\mathrm{Im}[\tilde{\mu}(\omega)]}{\omega P(\omega)}$ as a function of CW bias, where $P(\omega)$ is the power spectral density of the spontaneous fluctuations (Fig. 4a and Supplementary Figs 6 and 7). In the most general non-equilibrium case, the coupling coefficient K may change when the genetic background or the growth conditions are modified. In chemotaxis, we found that the value of the coupling coefficient $K(\omega)$ is independent of the functioning states of the cell and levels of expression of the chemotaxis proteins (Fig. 4a). This result is remarkable because most of the chemotaxis network has highly nonlinear signal processing[23,24].

It is usual to consider that noise is an independent limiting factor in intracellular signalling and that evolution selects network designs to reduce it[25]. However, using the framework of the FRT, we asked whether the temporal fluctuations in the switching rate of the motor and the cellular response are ever dynamically coupled. Remarkably, we found that the response time to a small external stimulus scaled linearly with the signalling noise from the chemotaxis network in cells before stimulus ($R^2 = 0.8$; Fig. 4b), which was consequently linearly related to the correlation time ($R^2 = 0.9$; Fig. 4c). Furthermore, we found that the response time to a large stimulus scaled poorly ($R^2 = 0.4$) with the signalling noise, reflecting that for large stimulus, the system operates outside the regime of linear approximation (Fig. 4b, inset).

We interpret this observation in simple mathematical terms, where the fluctuations in the network output, $\delta_{\text{CheY-P}}$, about its average have linearized kinetics in the form of a Langevin equation[21,26]: $\frac{\mathrm{d}}{\mathrm{d}t}\delta_{\text{CheY-P}} = -\frac{1}{\tau}\delta_{\text{CheY-P}} + \sqrt{D}\delta\eta(t)$, where $\sqrt{D}\delta\eta(t)$ is a white-noise source with intensity D and τ is the measured correlation time in the output of the signalling system. In this coarse-grained picture, there should exist a strict relationship between the signalling output noise amplitude $\sigma^2_{\text{CheY-P}}$ and the time τ, where $\sigma^2_{\text{CheY-P}} = (D/2)\tau$. Although the coefficient D could potentially depend on intracellular parameters in a complex way, our experiments surprisingly showed that two cellular traits, $\sigma^2_{\text{CheY-P}}$ and the response time, are linearly coupled. This observation implies that the coefficient D remains approximately constant over a wide range of functioning states of the cell (that is, CW bias). This result is consistent with the fact that the coefficient $\langle K(\omega)\rangle_\omega$ (Fig. 4a) determines the behaviour of D, because $\langle K(\omega)\rangle_\omega \propto 1/D$. Consequently, we anticipate that below an upper bound imposed mainly by rotational diffusion[27], cells with the largest behavioural variability before stimulus would also exhibit the strongest chemotactic drift in response to an external stimulus[21].

Although the FRT predicts the existence of a coupling between cellular response and noise, it does not specify how this coupling depends on the different states of the cell. Therefore, we hypothesize that the specific design of the signalling pathways could govern such interdependence. We find that a simple kinetic model and experimental data support this hypothesis (Supplementary Fig. 8): in chemotaxis, the value of the coefficient D is governed by the adaptation mechanism that uses the classic futile cycle[21] as a core module in which two antagonistic enzymes regulate the activity of the kinase-receptor complexes. Because the futile cycle is a design shared by a large class of signalling pathways[21,28,29], it raises the possibility that for these systems, noise and cellular response are coupled in a similar way. To gain general insights into the selection of a specific coupling, we should examine how certain classes of design and function of networks may constrain the behaviour of this interdependence[30].

METHODS SUMMARY

Response time. For each cell (whose behaviour is defined by a specific CW bias bin), the response time was measured from the time of stimulus through all successive averaged CCW intervals that were longer than the mean pre-stimulus CCW interval length. This mean was obtained by averaging together the CCW interval lengths chosen at random time points within the binary time series of the non-stimulated cell.

Correlation time. To determine the correlation time of the CCW sequences, we used serial correlation coefficients (Supplementary Fig. 4c) for the CCW interval lengths[19,20]. We converted the correlated number of sequences to the real correlation time lengths, including the half-length of the first uncorrelated CCW interval. To determine whether the sequences in each lag (the number of preceding CCW intervals) were correlated, we used the Wilcoxon rank sum test (the "ranksum" Matlab function) at a significance level of $P = 0.01$ (Supplementary Fig. 4d), as in ref. 20. We considered the first non-zero lag that had $h = 0$ as the end of the correlation.

Low-frequency noise and motor noise. We define the low-frequency noise N_i^{LF} of the ith cell as the integrated power density $P_i(f)$ of the binary time series from $f_i = 1/1{,}500\ \mathrm{s}^{-1}$ to $f_f = 1/10\ \mathrm{s}^{-1}$, which is $N_i^{\mathrm{LF}} \equiv \int_{f_i}^{f_f} P_i(f)\mathrm{d}f$ (Fig. 3a). We define the low-frequency motor noise $N_i^{\mathrm{LF,M}}$ as the integrated flat baseline of the power density (Fig. 3a, dark grey line) on the same timescale. We estimated signalling noise from the average experimental power spectral density, the average CW bias, and the gain function between the input signal (steady-state [CheY-P]) and output signal (average CW bias) using methods introduced by ref. 22 (Methods).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 June; accepted 4 October 2010.

Published online 14 November; corrected 9 December 2010 (see full-text HTML version for details).

1. Bourret, R. B., Borkovich, K. A. & Simon, M. I. Signal transduction pathways involving protein phosphorylation in prokaryotes. *Annu. Rev. Biochem.* **60,** 401–441 (1991).
2. Bourret, R. B. & Stock, A. M. Molecular information processing: lessons from bacterial chemotaxis. *J. Biol. Chem.* **277,** 9625–9628 (2002).
3. Korobkova, E. A., Emonet, T., Park, H. & Cluzel, P. Hidden stochastic nature of a single bacterial motor. *Phys. Rev. Lett.* **96,** 058105 (2006).
4. Callen, H. B. & Welton, T. A. Irreversibility and generalized noise. *Phys. Rev.* **83,** 34–40 (1951).
5. Prost, J., Joanny, J. F. & Parrondo, J. M. Generalized fluctuation-dissipation theorem for steady-state systems. *Phys. Rev. Lett.* **103,** 090601 (2009).
6. Bustamante, C., Macosko, J. C. & Wuite, G. J. L. Grabbing the cat by the tail: manipulating molecules one by one. *Nature Rev. Mol. Cell Biol.* **1,** 130–136 (2000).
7. Dorignac, J., Kalinowski, A., Erramilli, S. & Mohanty, P. Dynamical response of nanomechanical oscillators in immiscible viscous fluid for *in vitro* biomolecular recognition. *Phys. Rev. Lett.* **96,** 186105 (2006).
8. Paulsson, J. Summing up the noise in gene networks. *Nature* **427,** 415–418 (2004).
9. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genet.* **31,** 69–73 (2002).
10. Cugliandolo, L. F., Dean, D. S. & Kurchan, J. Fluctuation-dissipation theorems and entropy production in relaxational systems. *Phys. Rev. Lett.* **79,** 2168–2171 (1997).
11. Chetrite, R., Falkovich, G. & Gawedzki, K. Fluctuation relations in simple examples of non-equilibrium steady states. *J. Stat. Mech.-Theory E* **2008,** P08005 (2008).
12. Speck, T. & Seifert, U. Restoring a fluctuation-dissipation theorem in a nonequilibrium steady state. *Europhys. Lett.* **74,** 391–396 (2006).
13. Berg, H. C. Motile behavior of bacteria. *Phys. Today* **53,** 24–29 (2000).
14. Korobkova, E., Emonet, T., Vilar, J. M., Shimizu, T. S. & Cluzel, P. From molecular noise to behavioural variability in a single bacterium. *Nature* **428,** 574–578 (2004).
15. Cluzel, P., Surette, M. & Leibler, S. An ultrasensitive bacterial motor revealed by monitoring signaling proteins in single cells. *Science* **287,** 1652–1655 (2000).
16. Sourjik, V. & Berg, H. C. Receptor sensitivity in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA* **99,** 123–127 (2002).
17. Barkai, N. & Leibler, S. Robustness in simple biochemical networks. *Nature* **387,** 913–917 (1997).
18. Alon, U., Surette, M. G., Barkai, N. & Leibler, S. Robustness in bacterial chemotaxis. *Nature* **397,** 168–171 (1999).
19. Anderson, R. L. Distribution of the serial correlation coefficient. *Ann. Math. Stat.* **13,** 1–13 (1942).
20. Ratnam, R. & Nelson, M. E. Nonrenewal statistics of electrosensory afferent spike trains: implications for the detection of weak sensory signals. *J. Neurosci.* **20,** 6672–6683 (2000).
21. Emonet, T. & Cluzel, P. Relationship between cellular response and behavioral variability in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA* **105,** 3304–3309 (2008).
22. Shibata, T. & Fujimoto, K. Noisy signal amplification in ultrasensitive signal transduction. *Proc. Natl Acad. Sci. USA* **102,** 331–336 (2005).
23. Bray, D., Levin, M. D. & Morton-Firth, C. J. Receptor clustering as a cellular mechanism to control sensitivity. *Nature* **393,** 85–88 (1998).
24. Sourjik, V. & Berg, H. C. Functional interactions between receptors in bacterial chemotaxis. *Nature* **428,** 437–441 (2004).
25. Rao, C. V., Wolf, D. M. & Arkin, A. P. Control, exploitation and tolerance of intracellular noise. *Nature* **420,** 231–237 (2002).

26. Bialek, W. & Setayeshgar, S. Physical limits to biochemical signaling. *Proc. Natl Acad. Sci. USA* **102,** 10040–10045 (2005).
27. Andrews, B. W., Yi, T. M. & Iglesias, P. A. Optimal noise filtering in the chemotactic response of *Escherichia coli*. *PLOS Comput. Biol.* **2,** e154 (2006).
28. Goldbeter, A. & Koshland, D. E. Jr. An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl Acad. Sci. USA* **78,** 6840–6844 (1981).
29. Detwiler, P. B., Ramanathan, S., Sengupta, A. & Shraiman, B. I. Engineering aspects of enzymatic signal transduction: photoreceptors in the retina. *Biophys. J.* **79,** 2801–2817 (2000).
30. Shinar, G. & Feinberg, M. Structural sources of robustness in biochemical reaction networks. *Science* **327,** 1389–1391 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was funded by an NSF DMR award 0213745 to the Materials Research Science and Engineering Center at the University of Chicago, and NIH award R01AI059195-03 (to P.C.). W.P. and T.E. were supported by NSF CCF0829836, an Alfred P. Sloan Research Fellowship, and a National Academies Keck Futures Initiative award (to T.E.). J.F.M. was supported by NSF awards PHY-0852130 and DMR-0715099 and NIH grant 1U54CA143869-01. This work was also supported by the Chicago Biomedical Consortium with support from The Searle Funds at The Chicago Community Trust. D. Trentham supplied caged L-aspartate. We thank J. S. Parkinson for ΔCheB mutant strains RP4972 and RP4992. We thank T. Shimizu for discussions and sharing unpublished work. We thank H. Lee for help with the HPLC measurements. We thank J. Moffitt and K. Wood for comments on the manuscript and all members of the Cluzel laboratory for many discussions. W. Grus provided editorial assistance.

Author Contributions P.C. conceived and designed the research. H.P. performed all the experiments. H.P., P.C., T.E., W.P. and J.F.M. analysed the data. H.P., P.C., J.F.M. and T.E. wrote the paper. C.C.G. constructed the pZE21-CheR plasmid.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.C. (cluzel@mcb.harvard.edu).

METHODS

Strains and plasmids. RP437 is a wild-type *E. coli* strain for chemotaxis[31]. To construct pZE21-CheR, we amplified *cheR* using polymerase chain reaction (PCR) from the chromosome of the RP437 strain with the following primers: CheR-KpnI-5′: 5′-*gcc ggt acc atg act tca tca tct ctg ccc tg*-3′ and CheR-HindIII-3′: 5′-*cgc aag ctt tta atc ctt act tag cgc at*-3′. The gene fragment was inserted in the KpnI and HindIII sites of a pZE21 series plasmid[30] that contained a kanamycin resistance cassette and a TetR inducible promoter. The plasmid pZS4-Int1 encodes *tetR* under a constitutive promoter, which modulates the expression of the TetR-regulated *cheR* construct[32]. This plasmid carries a spectinomycin resistance gene. Wild-type cells with and without plasmid exhibited similar noise levels (Fig. 3a) and CCW interval lengths after stimulus (Supplementary Fig. 1a, c and d) at the single-cell level.

HPLC calibration of the release of aspartate. We prepared 10 μl samples of 0.5-mM caged L-aspartate solution under the same conditions for the chemotaxis experiments and illuminated them with intense ultraviolet light from a Xenon flash lamp (built-in L7685 reflective mirror, 60 W, Hamamatsu). We estimated the relative concentration of the caged L-aspartate in each sample by the high-performance liquid chromatography (HPLC) peak area. By comparing the decreasing HPLC peak area with its initial peak area, we found the released L-aspartate concentration as a function of the number of ultraviolet flashes (Supplementary Fig. 9). The samples released about 1 μM L-aspartate per ultraviolet flash. The HPLC gradient conditions had five steps: (1) equilibrium with 20% acetonitrile, 0.1% TFA/80% water, 0.1% TFA; (2) gradient of 20–55% acetonitrile over 30 min; (3) first washing with 55–90% acetonitrile for 20 min; (4) second washing with 90% acetonitrile for 5 min; and (5) equilibrium with 20% acetonitrile, 0.1% TFA/80% water, 0.1% TFA.

Photo-release and single-cell assay. We sheared the flagella of the cells by slowly forcing them through a thin needle (inner diameter = 0.19 mm, 27 G ½, B-D) 40 times. Cultures grew overnight in 3 ml of tryptone broth at 35 °C with shaking at 200 r.p.m. We transferred the overnight cultures to a 250 ml flask, in which we diluted them 1:50 in 12-ml tryptone broth and grew the cells again at 35 °C at 200 r.p.m. To obtain cells with different CW biases, we induced plasmid expression with various concentrations of anhydrotetracycline (0–2.5 ng ml^{-1}) in the diluted overnight cultures. The media also contained the antibiotic specific to the plasmid. We harvested the cells when the absorbance A reached ~0.3 at 600 nm. We washed the cells and resuspended them in motility medium (0.1 mM EDTA, 0.1 mM L-methionine, 10 mM potassium phosphate pH 7.0). We prepared glass slides (No. 1 ½, 18 mm, Corning) coated with poly-L-lysine and a solution of beads (Polybead Amino 1.0 μm Microspheres, Polysciences) coated with rabbit antibodies against flagella. We mixed the cells (4–5 μl) with the beads (4–5 μl) and incubated them for 20 min at room temperature (21–22 °C). This process caused the cell bodies to stick to the glass slide and the beads to attach to the flagella. Although the probability of a bead attaching to a rotating flagellum was low, we consistently obtained a few labelled flagella in each sample. After incubation we removed the unattached cells and beads and then added 8 μl of 5 μM (for small stimulus) or 500 μM (for large stimulus) caged L-aspartate solution to the sample medium. We covered the sample with oil (immersion oil transparent to ultraviolet: type FF, Cargille Laboratories) to prevent evaporation. We placed the sample under a dark-field condenser to produce a bright red image of the bead. Harmful blue light was filtered out by a long-pass filter (NT52-543, Edmund Industrial). We observed the samples under an Olympus IX71 microscope with an oil immersion objective 100× (numerical aperture = 1.3, Olympus Uplan FI, oil iris ∞/0.17). We recorded the long circular motions of individual beads attached to rotating flagella of single cells through a four-quadrant photomultiplier (type: R5900U-01-M4, Hamamatsu). The signal from the photomultiplier, a four-voltage time series, was monitored with a PC computer via LabView software (National Instrument). The rotation of the bead was simultaneously recorded using a charge-coupled device camera (1/3′′ midresolution Exview digital B/W camera, Sony). We converted the signal to a binary time series indicating transitions between CCW and CW rotations. After 1,500 s (or 300 s) of recording the rotational motion of the bead, we photo-released the caged aspartate (caged l-aspartic acid, sodium salt (189110): *N*-[1-(2-nitrophenyl)ethyloxycarbonyl]aspartic acid, sodium, $C_{13}H_{13}N_2O_8 \cdot Na$, relative molecular mass 348.2 and molar absorption $\varepsilon = 4{,}710\,M^{-1}\,cm^{-1}$ at maximum wavelength $\lambda_{max} = 264$ nm), from Calbiochem or synthesized by D. Trentham, G. Reid and J. Corrie). We illuminated the sample with an intense ultraviolet light from the Xenon flash coupled into a light guide (A2873, quartz glass fibre, Hamamatsu) and widely focused onto the whole sample with two ultraviolet-coated lenses (focal length = 35 mm and diameter = 25.4 mm; focal length = 20 mm and diameter = 12.7 mm, ThorLabs). These ultraviolet flashes produced a stepwise release of 1 μM (or 10 nM) L-aspartate from the 0.5 mM (or 5 μM) caged L-aspartate[33]. The magnitude of the stepwise stimulus corresponds to the typical increase in attractant concentration encountered by bacteria swimming in a gradient of 1 nM μm^{-1} (refs 34 and 35).

Definition of CW bias. We define $T_{i,j}^{CW}$ and $T_{i,j}^{CCW}$ as the durations of the jth CW and CCW intervals of the ith cell. The CW bias for the jth CW-to-CCW interval pair of the ith cell is $b_{i,j} = T_{i,j}^{CW} / \left(T_{i,j}^{CCW} + T_{i,j}^{CW}\right)$. The pre-stimulus CW bias of the ith cell, $\left\langle b_{i,j} \right\rangle_{before}$, is the time average of $b_{i,j}$ over a time window of length $t_{i,before}$ preceding the stimulus. $t_{i,\,before}$ was 300 s for the cells with CW bias exceeding 0.25 responding to the large stimulus and 1,500 s for all other cells. Similarly, the post-stimulus CW bias of the ith cell, $\left\langle b_{i,j} \right\rangle_{after}$, is the temporal average of $b_{i,j}$ over a time window of duration $t_{i,\,after}$ seconds following the stimulus. For the small (or large) stimulus, the first two (or 200) CW–CCW interval pairs following stimulus were not included. $t_{i,\,after}$ was 1,500 s for small stimuli, 900 s for large stimuli and CW bias <0.25, and 300 s for large stimuli and CW bias >0.25.

Response time. For each cell (the behaviour of which is defined by a specific CW bias bin), the response time was measured from the time of stimulus through all successive averaged CCW intervals that were longer than the mean pre-stimulus CCW interval length. This mean was obtained by averaging together the CCW interval lengths chosen at random time points within the binary time series of the non-stimulated cell. If the response time included more than one CCW interval, the CW interval length between two successive CCW intervals was also included in the response time. To get the final response time, we subtracted the mean non-stimulated portion of the first responding CCW interval. For example, if the third CCW interval is the last CCW interval length significantly longer than the mean CCW interval length before stimulus (dashed line in Figs 1b and c), the response time would be:

$$\langle T_{CCW,\,1st}\rangle + \langle T_{CW,\,1st}\rangle + \langle T_{CCW,\,2nd}\rangle + \langle T_{CW,\,2nd}\rangle + \langle T_{CCW,\,3rd}\rangle - \left\langle T_{CCW,\,1st,\,prestimulus}\right\rangle$$

The dashed line in Fig. 1b and c and Supplementary Fig. 1e represents the trend of the mean pre-stimulus CCW interval length in each CW bias bin. Because of the presence of a few outliers, we used the geometric mean to compute the trend of the mean CCW interval lengths after stimulus and mean pre-stimulus CCW interval length within each CW bias bin (Fig. 1b and c).

Correlation time. To determine the correlation time of the CCW sequences, we used serial correlation coefficients (Supplementary Fig. 4c) for the CCW interval lengths[19,20]. We converted the correlated number of sequences to the real correlation time lengths, including the half-length of the first uncorrelated CCW interval. To determine whether the sequences in each lag (the number of preceding CCW intervals) were correlated, we used the Wilcoxon rank sum test (the "ranksum" Matlab function) at a significance level of $P = 0.01$ (Supplementary Fig. 4d) as in ref. 20. We considered the first non-zero lag that had $h = 0$ as the end of the correlation.

Low-frequency noise and motor noise. We define the low frequency noise N_i^{LF} of the ith cell as the integrated power density $P_i(f)$ of the binary time series from $f_i = 1/1{,}500\,s^{-1}$ to $f_f = 1/10\,s^{-1}$, which is $N_i^{LF} \equiv \int_{f_i}^{f_f} P_i(f)df$ (Fig. 3a). We define the low-frequency motor noise $N_i^{LF,M}$ as the integrated flat baseline of the power density (Fig. 3a, dark grey line) on the same timescale.

Estimating signalling noise. To estimate the signalling noise, we used a formula $\sigma_{M,\,total}^2 \cong \sigma_M^2 + g_M{}^2\bar{b}^2 \dfrac{\sigma_{CheY\text{-}P}^2}{\overline{[CheY\text{-}P]}^2}$ which shows the relationship between the variance $\sigma_{CheY\text{-}P}^2$ of [CheY-P] and the variance $\sigma_{M,\,total}^2$ of the output signals. This formula was derived from a model recently introduced to describe generally the gain–noise relationship between the input and output signals in the chemical reaction network[22]. As ref. 22 showed, the temporally fluctuating output signal from a well defined steady state (CW bias = $\bar{b}$) due to the fluctuating input signal ([CheY-P]) is described by the following linearized chemical Langevin equation: $\dot{\delta b} = \gamma_M \delta[CheY\text{-}P] - \delta b/\tau_M + \xi_M(t)$, where δb and δ[CheY-P] are small deviations of the CW bias and [CheY-P] from their steady values, respectively, τ_M is the typical timescale of the motor alone and $\xi_M(t)$ is the Gaussian white-noise term that satisfies $\overline{\xi_M(t)} = 0$ and $\overline{\xi_M(t)\xi_M(t')} = \sigma_{\xi_M}^2 \cdot \delta(t - t')$. From this equation, we obtain the total variance of the output signals due to the temporally fluctuating input signals and the Gaussian white noise:

$$\sigma_{M,\,total}^2 = \frac{g_M \bar{b}}{\Theta_M} + g_M{}^2 \bar{b}^2 \frac{\tau_{CheY\text{-}P}}{\tau_M + \tau_{CheY\text{-}P}} \frac{\sigma_{CheY\text{-}P}^2}{\overline{[CheY\text{-}P]}^2}$$

where $\overline{[CheY\text{-}P]}$ is the steady value of fluctuating [CheY-P] values given by:

$$\overline{[CheY\text{-}P]} = K_M \left(\frac{\bar{b}}{1 - \bar{b}}\right)^{1/N_H}$$

where K_M (half the concentration of CheY-P that yields CW bias = 0.5) and the Hill coefficient N_H are given by 3.1 μM and 10.3, respectively, in ref. 15). The constant Θ_M in the first term is defined by $\Theta_M \equiv 2\gamma_M \overline{[\text{CheY-P}]} \Big/ {\sigma_{\xi_M}}^2$ and $\overline{b}$ is the CW bias. g_M is the gain function defined as the ratio of the fractional change of the output signal to the input signal: that is, $g_M = \left(\delta b/\overline{b}\right) \Big/ \left(\delta[\text{CheY-P}]/\overline{[\text{CheY-P}]}\right) = N_H\left(1-\overline{b}\right)$, where $N_H\left(1-\overline{b}\right)$ is obtained from ref. 15. $\tau_{\text{CheY-P}}$ is a characteristic timescale of the [CheY-P] fluctuations and is proportional to the input noise $\sigma^2_{\text{CheY-P}}$ as follows: $\tau_{\text{CheY-P}} = \frac{\sigma^2_{\xi_{\text{CheY-P}}}}{2}\sigma^2_{\text{CheY-P}}$. This relationship is derived from the chemical Langevin equation describing the [CheY-P] fluctuations from its steady state ($\overline{[\text{CheY-P}]}$):

$$\delta[\dot{\text{CheY}}\text{-P}] = -\frac{\delta[\text{CheY-P}]}{\tau_{\text{CheY-P}}} + \xi_{\text{CheY-P}}(t)$$

where $\xi_{\text{CheY-P}}(t)$ is a Gaussian white-noise term that satisfies $\overline{\xi_{\text{CheY-P}}(t)} = 0$ and $\overline{\xi_{\text{CheY-P}}(t) \cdot \xi_{\text{CheY-P}}(t')} = \sigma^2_{\xi_{\text{CheY-P}}} \cdot \delta(t-t')$. As long as the external stimulus is small enough, the response time to the stimulus should scale to $\tau_{\text{CheY-P}}$. For the broad range of the functioning states of this paper, we have one condition, $\tau_{\text{CheY-P}} \gg \tau_M$, in the timescales involved in this system. Under this condition, the above formula for the total variance of the output signals can be simplified to

$$\sigma^2_{\text{M, total}} \cong \sigma^2_M + {g_M}^2 \overline{b}^2 \frac{\sigma^2_{\text{CheY-P}}}{\overline{[\text{CheY-P}]}^2}$$

where $\sigma^2_{\text{M, total}}$ is given by $\overline{b}\left(1-\overline{b}\right)$ for any binary time series and is equal to the integral of the power spectral density over all frequencies (black line in Supplementary Fig. 5) averaged over all cells (wild-type RP437 and RP437 expressing CheR from pZE21-CheR) and σ^2_M is equal to the integral of the power density (dark grey line in Supplementary Fig. 5) of the isolated motor. We approximated the baseline of the motor power density by finding the mean value of the flat regime (from $f_i = 1/10\,\text{s}^{-1}$ to $f_f = 1/5\,\text{s}^{-1}$) of the average experimental power density and extending the baseline to the lowest frequency. By using the simplified formula above, we estimated the $\sigma^2_{\text{CheY-P}}$ values in each CW bias bin (Fig. 3b).

Definition of noise. We hypothesize that a small number of proteins and thermally activated biochemical reaction rates cause stochastic fluctuations between functional states of signalling proteins. Operationally, we monitor the cellular behaviour in a motility medium that does not support growth but allows bacteria to perform chemotaxis. Under these conditions, the observed noise does not result from protein synthesis or degradation; rather, it results from fluctuations in protein functional states about a well-defined steady state.

31. Parkinson, J. S. & Houts, S. E. Isolation and behavior of *Escherichia coli* deletion mutants lacking chemotaxis functions. *J. Bacteriol.* **151,** 106–113 (1982).
32. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25,** 1203–1210 (1997).
33. Jasuja, R., Yu-Lin, Trentham, D. R. & Khan, S. Response tuning in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA* **96,** 11346–11351 (1999).
34. Adler, J. A method for measuring chemotaxis and use of the method to determine optimum conditions for chemotaxis by *Escherichia coli*. *J. Gen. Microbiol.* **74,** 77–91 (1973).
35. Bainer, R., Park, H. & Cluzel, P. A high-throughput capillary assay for bacterial chemotaxis. *J. Microbiol. Methods* **55,** 315–319 (2003).

doi:10.1038/nature09557

Tumour vascularization via endothelial differentiation of glioblastoma stem-like cells

Lucia Ricci-Vitiani[1]*, Roberto Pallini[2]*, Mauro Biffoni[1], Matilde Todaro[3], Gloria Invernici[4], Tonia Cenci[5], Giulio Maira[2], Eugenio Agostino Parati[4], Giorgio Stassi[3,6], Luigi Maria Larocca[5] & Ruggero De Maria[1,7]

Glioblastoma is a highly angiogenetic malignancy, the neoformed vessels of which are thought to arise by sprouting of pre-existing brain capillaries. The recent demonstration that a population of glioblastoma stem-like cells (GSCs) maintains glioblastomas[1,2] indicates that the progeny of these cells may not be confined to the neural lineage[3]. Normal neural stem cells are able to differentiate into functional endothelial cells[4]. The connection between neural stem cells and the endothelial compartment seems to be critical in glioblastoma, where cancer stem cells closely interact with the vascular niche and promote angiogenesis through the release of vascular endothelial growth factor (VEGF) and stromal-derived factor 1 (refs 5–9). Here we show that a variable number (range 20–90%, mean 60.7%) of endothelial cells in glioblastoma carry the same genomic alteration as tumour cells, indicating that a significant portion of the vascular endothelium has a neoplastic origin. The vascular endothelium contained a subset of tumorigenic cells that produced highly vascularized anaplastic tumours with areas of vasculogenic mimicry in immunocompromised mice. *In vitro* culture of GSCs in endothelial conditions generated progeny with phenotypic and functional features of endothelial cells. Likewise, orthotopic or subcutaneous injection of GSCs in immunocompromised mice produced tumour xenografts, the vessels of which were primarily composed of human endothelial cells. Selective targeting of endothelial cells generated by GSCs in mouse xenografts resulted in tumour reduction and degeneration, indicating the functional relevance of the GSC-derived endothelial vessels. These findings describe a new mechanism for tumour vasculogenesis and may explain the presence of cancer-derived endothelial-like cells in several malignancies.

From archival material, we selected a group of glioblastomas showing both remarkable angiogenesis and nuclear accumulation of mutant p53 in tumour cells (Supplementary Table 1). In 83.3% (20/24) of these tumours, we found cells with nuclear accumulation of mutant p53 that lined the lumens of capillaries and/or vascular glomeruli (Supplementary Fig. 1a and Supplementary Table 1). Double immunohistochemistry analysis of p53 and CD31 demostrated the endothelial phenotype of the p53-positive cells facing the lumen of the vessels (Supplementary Fig. 1b). Mouse and human tumour-associated endothelial cells can harbour chromosomal alterations[10–12]. To assess whether a subset of endothelial cells showed glioblastoma-specific chromosomal aberrations, we analysed the tumour vasculature in 15 glioblastomas by combined CD31 immunofluorescence and fluorescence *in situ* hybridization (FISH) using probes for the centromere of chromosome 10 (Cep10), for the telomere of chromosome 19 (Tel19q), and a locus-specific probe on chromosome 22 (breakpoint cluster region locus q11.2; LSI22). In all the tumours carrying aneuploidy for one or more of these chromosomes, we detected a substantial fraction of endothelial cells bearing the same chromosomal aberrations (Supplementary Fig. 1c). Interestingly, double immunostaining of vascular glomeruli in glioblastoma revealed a significant number of $GFAP^+$ microvascular cells showing an aberrant endothelial/glial phenotype (Supplementary Fig. 1d). Thus, a variable number of endothelial cells in glioblastoma seem to originate from the tumour. To quantify the contribution of tumour-derived endothelial cells to glioblastoma vasculature, we used FISH to analyse purified $CD31^+/CD144^+$ (VE-Cadherin$^+$) endothelial cells from freshly dissociated glioblastoma specimens (Fig. 1a). Again, we detected $CD31^+/CD144^+$ endothelial cells that shared the same chromosomal alterations as the tumour cells in any given glioblastoma harbouring aberrations of chromosomes 10, 19 and 22 (Fig. 1b and Supplementary Table 2). The amount of endothelial cells with tumour-specific chromosomal changes ranged between 20 and 90% of the sorted cells (mean 60.7 ± 28.1 standard deviations (s.d.)).

We assessed further the phenotype of sorted $CD31^+/CD144^+$ glioblastoma cells by immunofluorescence, which showed that the vast majority of these cells ($83.9 \pm 4.2\%$; range 79–90%) expressed the mature endothelial cell marker von Willebrand factor (vWF), although a substantial proportion of them ($30.9 \pm 21.3\%$; range 10–76%) co-expressed vWF and GFAP (Fig. 1c, d). Thus, it seems that the $CD31^+/CD144^+$ cells harbouring chromosomal aberrations are glioblastoma-derived endothelial cells that either differentiated towards the canonical endothelial lineage ($GFAP^-$) or showed a mixed endothelial/glial phenotype ($GFAP^+$), whereas the euploid fraction is likely to represent endothelial cells derived from normal brain vessels. *In vitro* experiments using a microvascular culture of fresh $CD31^+$ cells isolated by magnetic microbeads from glioblastoma samples confirmed the existence of endothelial cells with aberrant GFAP expression (Supplementary Fig. 2a–c), as well as the presence of a substantial number of aneuploid endothelial cells (Supplementary Fig. 2d, e). Grafting of freshly purified $CD31^+/CD144^+$ cells showed that three of five glioblastomas contained tumorigenic endothelial cells that produced highly vascularized anaplastic tumours (Supplementary Fig. 3a–c). These cells, however, lost their tumorigenic activity on *in vitro* culture with endothelial medium (Supplementary Fig. 3a).

Although there is no general agreement on the definiton and markers identifying so-called cancer stem cells, there is good evidence that GSCs can be enriched by the use of anti-CD133 antibodies or through the generation of clusters of undifferentiated cells (neurospheres) in serum-free media containing epidermal growth factor (EGF) and basic fibroblast growth factor (FGF)[1,5,7,13,14]. We demonstrated recently that GSCs can differentiate into mesenchymal cells, giving rise to osteoblastic and chondrocytic cells[3]. To determine the potential contribution to the angiogenic process of GSCs, we cultivated glioblastoma neurospheres and primary glioblastoma differentiated cells under endothelial conditions, or $CD133^+/CD31^-$ and $CD133^-/CD31^-$ cells derived from the

[1]Department of Hematology, Oncology and Molecular Medicine, Istituto Superiore di Sanità, Viale Regina Elena 299, Rome 00161, Italy. [2]Department of Neurosurgery, Catholic University of Rome, Largo Francesco Vito 1, Rome 00168, Italy. [3]Department of Surgical and Oncological Sciences, University of Palermo, Via Liborio Giuffrè 5, Palermo 90127, Italy. [4]U. O. Cerebrovascular Disease Cellular Neurobiology Laboratory, Fondazione IRCCS, Neurological Institute ''Carlo Besta'', Via Celoria 11, 20133 Milan, Italy. [5]Institute of Pathology, Catholic University of Rome, Largo Francesco Vito 1, Rome 00168, Italy. [6]Cellular and Molecular Oncology, Istituto di Ricovero e Cura a Carattere Scientifico Fondazione Salvatore Maugeri, Via S. Maugeri 6, Pavia 27100, Italy. [7]Mediterranean Institute of Oncology, Via Penninazzo 7, 95029, Viagrande, Catania, Italy.
*These two authors contributed equally to the manuscript.

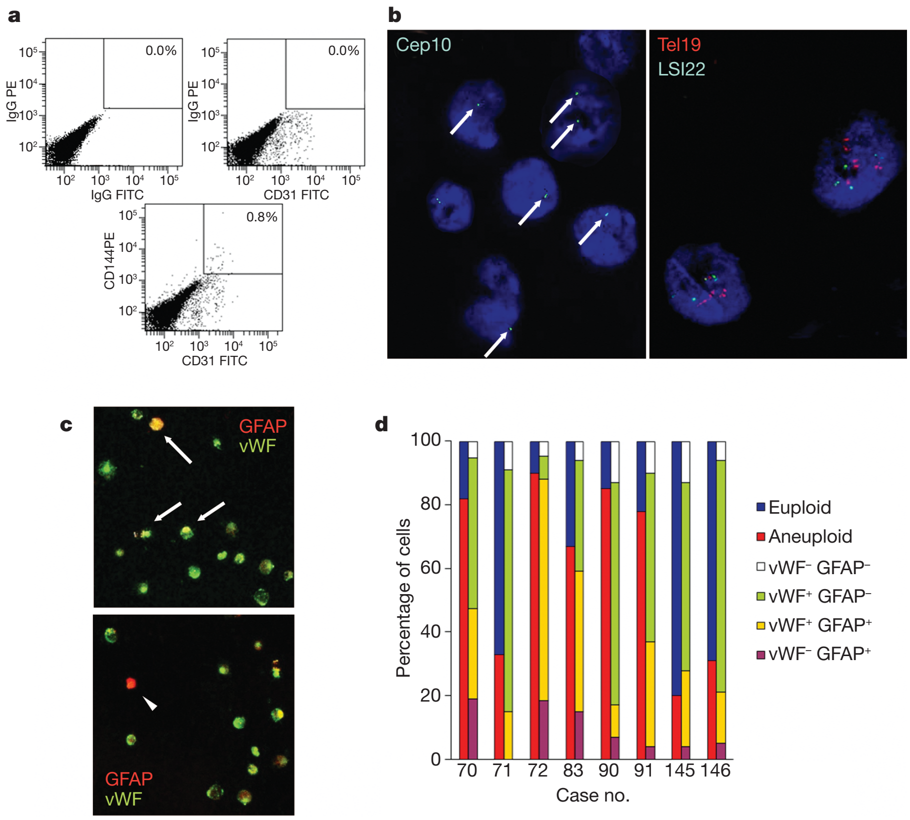


Figure 1 | Microvascular endothelial cells isolated from glioblastoma harbour tumour-specific chromosomal aberrations. **a**, CD31$^+$/CD144$^+$ cells were isolated from surgical glioblastoma specimens ($n = 15$). FITC, fluorescein isothiocyanate; PE, phycoerythrin. **b**, Sorted cells were analysed by interphase FISH assay for tumour-specific chromosomal changes, such as monosomy of Cep10 (left, arrows) or polisomy of Tel19 and LSI22 (right). **c**, The phenotype of the CD31$^+$/CD144$^+$ sorted cells was further analysed by anti-GFAP and anti-vWF immunofluorescence. A fraction of the CD31$^+$/CD144$^+$ cells coexpressed GFAP and vWF (arrows), indicating an aberrant endothelial/glial phenotype. A minority of sorted cells were GFAP$^+$/vWF$^-$. **d**, Quantification of results from FISH and immunofluorescence analysis.

same tumours. Whereas cells enriched in GSCs generated microvascular cultures of CD31$^+$ and Tie2$^+$ cells, neither differentiated cells nor the U87MG cell line were able to produce endothelial-like cells (Fig. 2a). Such GSC-derived endothelial cells showed considerable tube-forming ability, together with low-density lipoprotein (LDL) uptake and endothelial nitric oxide synthase (eNOS) expression, which were completely absent in differentiated tumour cells and in the U87MG cell line (Fig. 2b–d and Supplementary Fig. 4). Unsupervised gene-expression analysis of glioblastoma and endothelial cells showed that neural-differentiated glioblastoma cells and normal endothelial cells constitute the two more distant groups in a dendrogram in which tumour endothelial cells cluster between normal endothelial cells and glioblastoma neurospheres (Supplementary Fig. 5).

To investigate the ability of GSCs to form endothelial vessels *in vivo*, we measured the relative amount of murine versus human endothelial cells within glioblastoma neurosphere xenografts (Fig. 3a). Flow cytometry analysis with human- and mouse-specific antibodies showed that about 70% of the CD31$^+$ cells from the inner portion of the tumour were of human origin, whereas nearly all the CD31$^+$ cells in the tumour capsule were murine (Fig. 3b). Likewise, human CD144$^+$ cells were detected only in the core and not in the tumour capsule (Fig. 3b). Immunohistochemistry of subcutaneous and intracranial xenografts showed that glioblastoma neurosphere-derived tumours contained human vessels labelled by human-specific anti-CD31, whereas xenografts generated with U87MG or other glioma cell lines grown in serum did not (Fig. 3c, Supplementary Fig. 6a and data not shown). The presence of human-derived endothelial cells was confirmed by labelling sections of tumour xenografts obtained with GFP$^+$ glioblastoma neurospheres with anti-GFP and anti-human CD31 antibodies (Supplementary Fig. 6b). Moreover, immunofluorescence staining with validated human-specific endothelial antibodies showed that these cells expressed consistently CD34, CD144 and VEGFR2 (Fig. 3d) but not the stem-cell markers SSEA-1 and CD133 (Supplementary Fig. 6c). Such human-specific endothelial antigens identified microvascular structures containing circulating erythrocytes (Fig. 3d and Supplementary Fig. 6a), indicating the functional relevance of human angiogenesis in the tumour xenografts. Of note, a similar formation of human endothelial cells was observed in subcutaneous xenografts obtained with the injection of freshly purified CD133$^+$/CD31$^-$ cells, whereas CD133$^-$/CD31$^-$ cell xenografts contained only mouse endothelial vessels (Supplementary Fig. 7).

To trace *in vivo* angiogenesis, we injected RFP-labelled glioblastoma neurospheres into transgenic NOD/SCID mice expressing GFP under the Tie2 promoter. Examination of a thick-section plane by confocal microscopy showed that GFP$^+$ mouse vessels were primarily outside the tumours (Supplementary Fig. 8a). To exclude the occurrence of fusion between tumour and mouse endothelial cells, we stained tumour xenograft sections with anti-human/mouse Tie2 and CD31 antibodies. Although CD31 staining showed the presence of vessels containing both mouse and human CD31$^+$ cells at the periphery of the tumour, the majority of endothelial cells inside the tumour mass did not express mouse Tie2 and were of human origin in the absence of fusion (Supplementary Fig. 8b, c). Moreover, FISH analysis of nuclei extracted from microdissected vascular structures of GSC xenografts confirmed the absence of murine chromosomes in human cells (Supplementary Fig. 9). Together, these findings demonstrate that the tumour xenografts obtained by injection of human glioblastoma neurospheres develop an intrinsic vascular network composed by

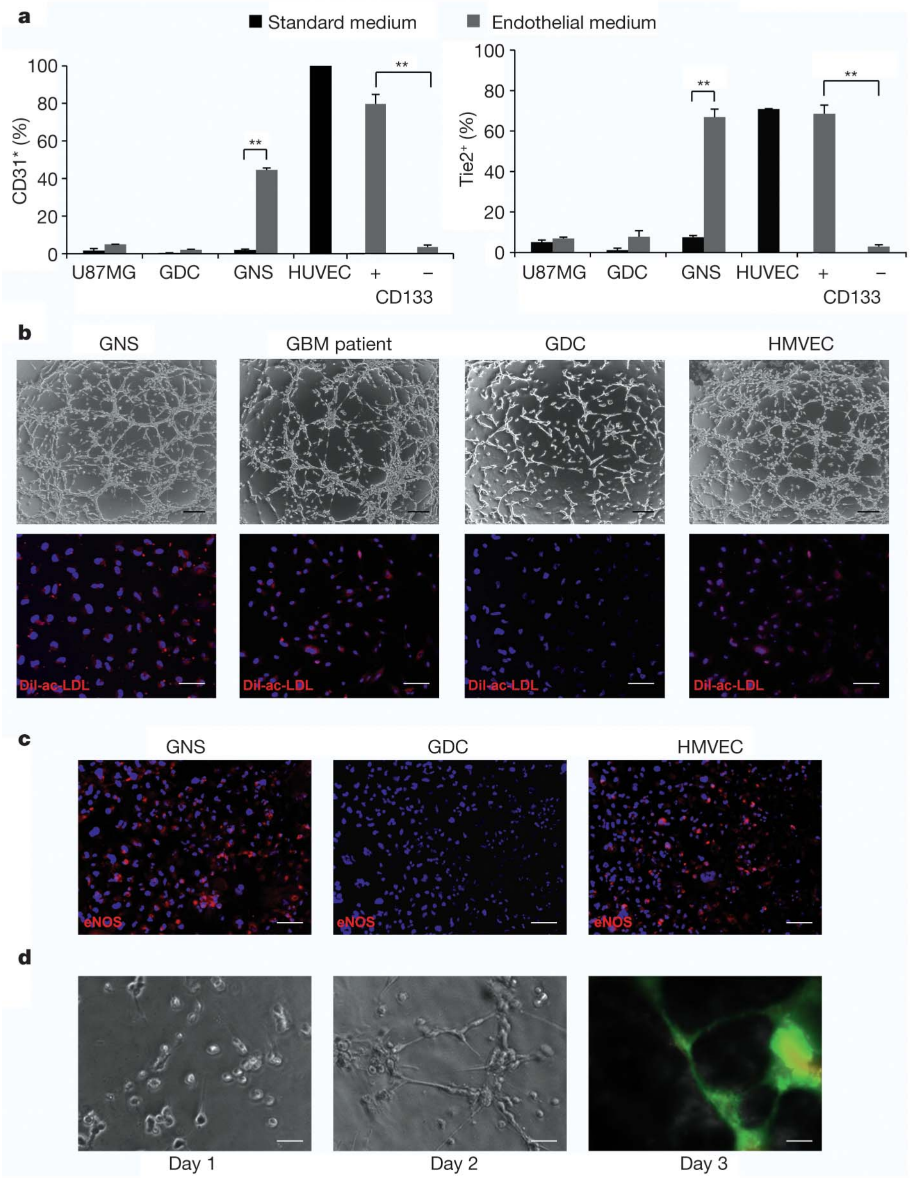


Figure 2 | GSCs cultured under endothelial differentiation conditions develop morphological, phenotypical and functional features of endothelial cells. a, Flow cytometry analysis of human umbilical vein endothelial cells (HUVEC), glioblastoma neurospheres (GNS), primary glioblastoma cells cultured in serum (GDC), U87MG, CD31^{-}/CD133^{+} and CD31^{-}/CD133^{-} cells from freshly dissociated glioblastomas. Cells were cultured under standard (black) or endothelial (grey) condition. Error bars represent the mean ± s.d. ($n = 4$). $^{**}P < 0.001$. **b**, Tube formation (top) and LDL-uptake (bottom) assay on cells under endothelial conditions as above (GNS and GDC), endothelial cells isolated from glioblastoma patients (GBM patients) and human dermal microvascular endothelial cells (HMVEC). DiI-ac-LDL, 1,1′-dioetadeeyl-3,3,3′,3′-tetramethylindocarboeyanine-perchlorate-acetylated LDL. Scale bars, 200 μm (top) and 50 μm (bottom). **c**, Immunofluorescence for eNOS in HMVEC, GNS and GDCs treated as above. Scale bar, 100 μm. **d**, *In vitro* perfusion assay on three-dimensional glioblastoma neurosphere-derived endothelial culture injected with fluorescein. Scale bar, 50 μm. One representative of four independent experiments performed in blind is shown for **b**, **c** and **d**.

tumour cells with an aberrant endothelial phenotype. To determine whether the GSC-derived endothelial cells contribute to tumour growth, we transduced glioblastoma neurospheres with a lentiviral vector containing the herpes simplex virus thymidine kinase gene (*tk*) under the control of the transcription-regulatory elements of Tie2 (Tie2-*tk*; Supplementary Fig. 10a), so that the tumour-derived endothelial cells would be sensitive to ganciclovir[15,16]. For this experiment, we selected glioblastoma neurospheres with no detectable expression of Tie2 (Supplementary Fig. 11). Control cells included glioblastoma neurospheres tranduced with an empty viral vector and U87MG cells transduced either with Tie2-*tk* or with a vector conferring constitutive expression of Tk (PGK-*tk*, Supplementary Fig. 10a). One week after ganciclovir administration, TdT-mediated dUTP nick end labelling (TUNEL) and double immunofluorescence labelling with anti-Tie2 antibodies in tumour subcutaneous xenografts showed selective apoptosis of the endothelial compartment only in animals injected with Tie2-*tk* neurospheres, whereas PGK-*tk* tumours contained a considerable number of apoptotic nuclei both in tumour and endothelial cells (Fig. 4a). Moreover, tumours generated by Tie2-*tk* neurospheres underwent a significant size reduction four weeks after ganciclovir administration, whereas control GSC xenografts increased their size over the same time interval (Fig. 4b).

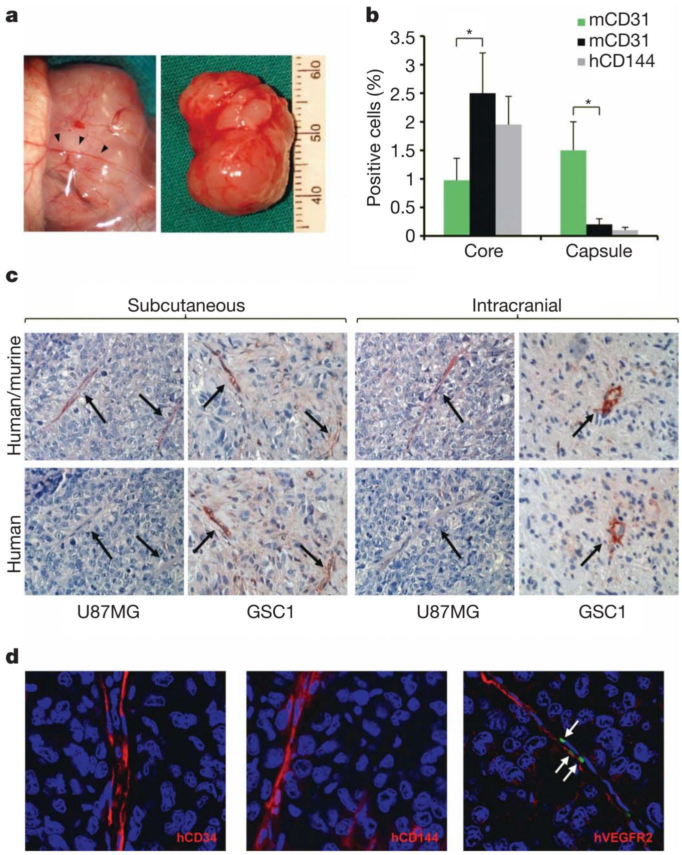


Figure 3 | Human origin of endothelial cells in glioblastoma neurosphere xenografts. a, Explanted subcutaneous xenograft obtained by injection of glioblastoma neurospheres. Detail of murine vessels on the surface of the xenograft (left, black arrowheads) and tumour after capsule removal (right). **b**, FACS evaluation of murine $CD31^{+}/CD45^{-}$ (mCD31), human CD31 (hCD31) and human CD144 (hCD144) in the capsule and core of the tumour (mean ± s.d., $n = 4$, $*P < 0.05$). **c**, Immunohistochemistry of glioblastoma neurosphere (GSC1) and U87MG xenografts using either an anti-human CD31 or anti-human and murine CD31 (one out of four different glioblastoma neurosphere samples and serum-grown cell lines are shown). **d**, Immunofluorescence of tumour xenograft sections labelled with anti-human CD34 (left), anti-human CD144 (middle) or anti-human VEGFR2 (right). Arrows indicate circulating erythrocytes. Data represent one of four independent experiments obtained with different glioblastoma neurosphere samples.

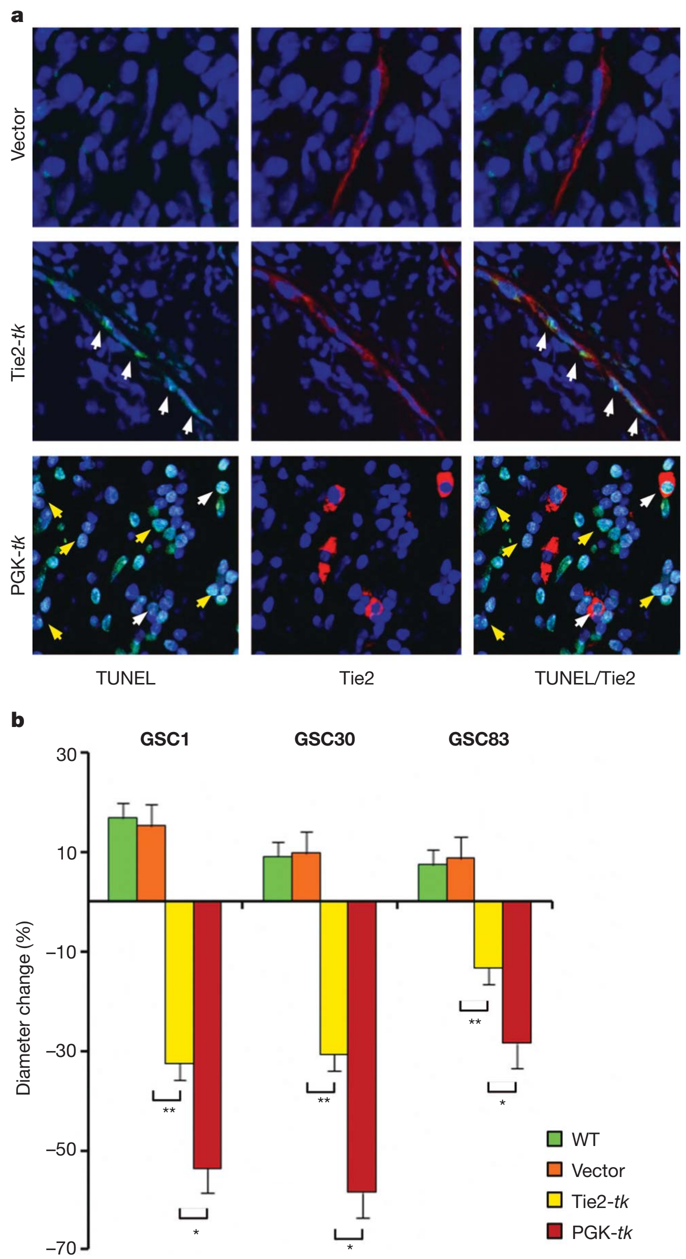


Figure 4 | Selective targeting of glioblastoma neurosphere-derived endothelial cells impairs the growth of subcutaneous tumour xenografts. a, Double immunofluorescence using anti-TUNEL and anti-Tie2 in xenografts from Tie2-*tk*, PGK-*tk* and vector glioblastoma neurosphere cells one week after ganciclovir administration. Arrows indicate apoptotic $Tie2^{+}$ (white) and $Tie2^{-}$ (yellow) cells. **b**, Tumour size measured four weeks after ganciclovir administration in xenograft obtained from three different glioblastoma neurosphere samples either untransduced (wild type (WT)) or transduced with vector, Tie2-*tk* or PGK-*tk*. Error bars are mean ± s.d. of three different experiments. $*P < 0.005$, $**P < 0.001$.

Histological examination revealed massive degeneration in the tumour xenografts developed by injection of Tie2-*tk* neurospheres. Four weeks after ganciclovir treatment, these tumours were completely devoid of vascular glomeruli, tiny capillaries with ongoing phenomena of endothelial disruption being the only residual vascular structures (Supplementary Fig. 10b). Although all PGK-*tk* tumours degenerated massively, U87MG Tie2-*tk* xenografts were not affected by ganciclovir treatment (Supplementary Fig. 10c, d), confirming that this cell line was unable to generate endothelial cells. These findings indicate that GSC-derived angiogenesis is essential for tumour survival. Moreover, mouse models based on adherent cell lines grown in serum do not seem suitable for the study of glioblastoma angiogenesis.

Here we demonstrated that GSCs are able to differentiate in functional endothelial cells. Such angiogenic potential could be inherited from normal neural stem cells, which have been shown to differentiate in endothelial cells both *in vitro* and *in vivo*[4]. The formation of fluid-conducting networks by nonendothelial cells has been described for melanomas, sarcomas, breast, ovary, lung and prostate carcinomas[17,18] as a result of vasculogenic mimicry, which is a feature associated with a pluripotent gene expression pattern in aggressive tumour cells[19].

The ability of cancer stem-like cells to directly contribute to the tumour vasculature by endothelial cell differentiation represents a new mechanism of angiogenesis that might not be restricted to glioblastoma. A similar endothelial potential may be shared by $CD44^{+}$ cells purified from ovarian cancer[20]. However, the existence of tumour-derived endothelial cells in ovarian cancer has not been demonstrated yet. Endothelial-like cells with cancer-specific genomic alterations have been described in other tumour types, such as lymphoma and neuroblastoma[11,12]. Although the angiogenic activity of cancer stem-like cells has not been investigated in other tumours, it is likely that the

endothelial cells bearing tumour-specific alterations derive from cancer cells endowed with stem-cell plasticity. Likewise, the vasculogenic mimicry might represent an incomplete differentiation of cancer stem-like cells towards the endothelial lineage, as indicated by the aberrant mixed phenotype of glioblastoma xenografts generated by the subset of $CD31^+/CD144^+$ cells that retain tumorigenic activity.

Our findings may have considerable therapeutic implications. On the one hand, endothelial cells bearing the same genomic alteration as cancer cells may show a different sensitivity to conventional anti-angiogenic treatments, such as VEGF/VEGFR targeting. On the other hand, our data indicate the possibility of targeting the process of GSC differentiation into endothelial cells, thus offering new therapeutic options for cancer treatment.

METHODS SUMMARY

Cell culture. Glioblastoma neurosphere cultures were established from freshly dissociated surgical specimens as described[3,21,22]. Primary cultures of glioblastoma differentiated cells were obtained by plating cells from freshly dissociated samples in DMEM-F12 medium containing 10% FBS. For primary culture of glioblastoma microvascular endothelial cells, $CD31^+$ cells were purified using Miltenyi Microbead Kit (Miltenyi Biotech) according to manufacturer's instructions and grown in endothelial basal medium (EBM Bullet kit; Biowhittaker Cambrex).

Immunohistochemistry, immunofluorescence and flow cytometry. Immunohistochemistry was performed as described[22] on deparaffinized sections of glioblastoma tissue. For immunofluorescence, cells were fixed with 4% paraformaldehyde and permeabilized in 0.1% Triton X-100. Cytofluorimetric analysis was performed using a FACS Canto flow cytometer (Becton Dickinson). Cell sorting was performed with a FACS Aria cell sorter (Becton Dickinson).

Interphase FISH and combined immunohistochemistry and FISH (FICTION). Single- and dual-probe interphase FISH was performed as described[3]. Images were captured using a high-resolution black and white CCD microscope camera AxioCam MRm REV 2 (Karl Zeiss) and analysed using AxioVision 4 multichannel fluorescence basic workstation (Karl Zeiss).

Lentiviral infection. Selective targeting of the cells expressing endothelial phenotype was obtained by modifying the pRRLsin.Tie2p.TKiresGFP.spre lentiviral vector provided by L. Naldini[15,16]. Viral particle production and GSC infection were performed as previously described[23].

***In vivo* experiments.** Nude athymic and SCID mice (female, 4–5 weeks of age; Charles River) were used. Partially dissociated glioblastoma neurospheres were used for both orthotopic and subcutaneous injection, typically 10^5 and 5×10^5, respectively. For *in vivo* endothelial targeting, mice were injected with Tie2-*tk* glioblastoma neurospheres into the right flank and control vector glioblastoma neurospheres into the left flank. After having developed bilateral nodules mice received ganciclovir at 50 mg kg^{-1} day^{-1} intraperitoneally for 5 days. Ganciclovir-treated mice were killed at different time points to collect samples for histology and immunofluorescence.

Statistical analysis. Student's *t*-test was used to analyse data using Statistica (version 5.5; Statsoft) or Fig.P (version 2.7; Biosoft) softwares.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 October 2009; accepted 13 September 2010.
Published online 21 November 2010.

1. Singh, S. K. *et al.* Identification of human brain tumour initiating cells. *Nature* **432,** 396–401 (2004).
2. Vescovi, A. L., Galli, R. & Reynolds, B. A. Brain tumour stem cells. *Nature Rev. Cancer* **6,** 425–436 (2006).
3. Ricci-Vitiani, L. *et al.* Mesenchymal differentiation of glioblastoma stem cells. *Cell Death Differ.* **15,** 1491–1498 (2008).
4. Wurmser, A. E. *et al.* Cell fusion-independent differentiation of neural stem cells to the endothelial lineage. *Nature* **430,** 350–356 (2004).
5. Bao, S. *et al.* Stem cell-like glioma cells promote tumor angiogenesis through vascular endothelial growth factor. *Cancer Res.* **66,** 7843–7848 (2006).
6. Folkins, C. *et al.* Glioma tumor stem-like cells promote tumor angiogenesis and vasculogenesis via vascular endothelial growth factor and stromal-derived factor 1. *Cancer Res.* **69,** 7243–7251 (2009).
7. Calabrese, C. *et al.* A perivascular niche for brain tumor stem cells. *Cancer Cell* **11,** 69–82 (2007).
8. Gilbertson, R. J. & Rich, J. N. Making a tumour's bed: glioblastoma stem cells and the vascular niche. *Nature Rev. Cancer* **7,** 733–736 (2007).
9. Lathia, J. D. *et al.* Integrin α6 regulates glioblastoma stem cells. *Cell Stem Cell* **6,** 421–432 (2010).
10. Hida, K. *et al.* Tumor-associated endothelial cells with cytogenetic abnormalities. *Cancer Res.* **64,** 8249–8255 (2004).
11. Streubel, B. *et al.* Lymphoma-specific genetic aberrations in microvascular endothelial cells in B-cell lymphomas. *N. Engl. J. Med.* **351,** 250–259 (2004).
12. Pezzolo, A. *et al.* Tumor origin of endothelial cells in human neuroblastoma. *J. Clin. Oncol.* **25,** 376–383 (2007).
13. Singh, S. K. *et al.* Identification of a cancer stem cell in human brain tumors. *Cancer Res.* **63,** 5821–5828 (2003).
14. Galli, R. *et al.* Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. *Cancer Res.* **64,** 7011–7021 (2004).
15. De Palma, M., Venneri, M. A. & Naldini, L. *In vivo* targeting of tumor endothelial cells by systemic delivery of lentiviral vectors. *Hum. Gene Ther.* **14,** 1193–1206 (2003).
16. De Palma, M., Venneri, M. A., Roca, C. & Naldini, L. Targeting exogenous genes to tumor angiogenesis by transplantation of genetically modified hematopoietic stem cells. *Nature Med.* **9,** 789–795 (2003).
17. Hendrix, M. J., Seftor, E. A., Hess, A. R. & Seftor, R. E. Vasculogenic mimicry and tumour-cell plasticity: lessons from melanoma. *Nature Rev. Cancer* **3,** 411–421 (2003).
18. Dome, B., Hendrix, M. J., Paku, S., Tovari, J. & Timar, J. Alternative vascularization mechanisms in cancer: pathology and therapeutic implications. *Am. J. Pathol.* **170,** 1–15 (2007).
19. Maniotis, A. J. *et al.* Vascular channel formation by human melanoma cells *in vivo* and *in vitro*: vasculogenic mimicry. *Am. J. Pathol.* **155,** 739–752 (1999).
20. Alvero, A. B. *et al.* Molecular phenotyping of human ovarian cancer stem cells unravels the mechanisms for repair and chemoresistance. *Cell Cycle* **8,** 158–166 (2009).
21. Eramo, A. *et al.* Chemotherapy resistance of glioblastoma stem cells. *Cell Death Differ.* **13,** 1238–1241 (2006).
22. Pallini, R. *et al.* Cancer stem cell analysis and clinical outcome in patients with glioblastoma multiforme. *Clin. Cancer Res.* **14,** 8205–8212 (2008).
23. Ricci-Vitiani, L. *et al.* Absence of caspase 8 and high expression of PED protect primitive neural cells from cell death. *J. Exp. Med.* **200,** 1257–1266 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Naldini for providing the lentiviral vectors and S. Forte for gene array data elaboration. This work was supported by grants from Associazione Italiana per la Ricerca sul Cancro.

Author Contributions L.R.-V. and R.P. performed most of the experiments and coordinated the project; M.B. performed cell sorting and flow cytometric analysis; M.T. and G.S. detected and characterized human endothelial cells in mouse xenografts; G.I. and E.A.P. developed the functional assays of the endothelial cell cultures; G.M. recruited the patients and performed surgery; T.C. and L.M.L. were involved in pathology assessment and detection of genomic aberration in endothelial cells; R.D.M. conceived the study and wrote the paper.

Author Information Data have been deposited at the European Bioinformatics Institute (http://www.ebi.ac.uk/arrayexpress/) under accession number E-MEXP-2891. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.D.M. (demaria@iss.it) or R.P. (pallini@rm.unicatt.it).

METHODS

Glioblastoma neurosphere isolation and characterization. Glioblastoma tissue specimens were obtained from adult patients undergoing craniotomy at the Institute of Neurosurgery, Catholic University School of Medicine in Rome. Informed consent was obtained before surgery according to the protocols approved at the Catholic University. Cells were purified through mechanical dissociation of the tumour tissue and cultured at clonal density in a serum-free medium supplemented with EGF and basic FGF as described[3,21,22]. Isolated cells were expanded and characterized both *in vitro* and *in vivo*. In these conditions, cells were able to grow *in vitro* in clusters called neurospheres and maintain an undifferentiated state, as indicated by morphology and expression of stem-cell markers such as CD133, SOX2, musashi and nestin. Such glioma neurosphere cells showed a clonal frequency higher than 10%, ability to coexpress astrocytic as well as neuronal phenotypic markers after serum-induced differentiation *in vitro*, and generation of glial tumours in immunodeficient mice.

Flow cytometry, immunohistochemistry and immunofluorescence. Cell suspension obtained by mechanical dissociation of the tumour tissue from glioblastoma patients from the Institute of Neurosurgery (Supplementary Table 2) was passed through a 100-μm mesh to remove aggregates and stained with fluorochrome-conjugated antibodies to surface antigens. After 1h of incubation on ice, cells were washed twice with PBS and finally resuspended in PBS or in PBS containing 7-aminoactinomycin D (7-AAD) 5μg ml^{-1} to assess viability. Analysis was performed using a fluorescence-activated cell sorter (FACS) Canto flow cytometer (Becton Dickinson). Cell sorting was performed with a FACS Aria cell sorter (Becton Dickinson) equipped with an automatic cloning deposition unit. Cells were selected on the basis of physical parameters and fluorescence and were sorted on sterile tubes or on slides depending on their further utilization.

For immunoistochemistry, immunofluorescence and flow cytometry the following antibodies were used: mouse anti-human CD31 (Novocastra); mouse anti-CD31 (Dako); rat anti-mouse CD31 (BD, Pharmingen); rabbit anti-GFAP (Chemicon or Dako); mouse anti-vWF (Dako); goat anti-human Tie2 (R&D Systems); mouse anti-human VEGFR2 (R&D Systems); rabbit anti-Tie2 (Santa Cruz Biotechnology); mouse anti-human CD144 (R&D Systems); mouse anti-SSEA-1 (R&D Systems); mouse anti-human nuclei antigen (Chemicon); rabbit anti-GFP (Molecular Probes) and anti-eNOS (BD, Pharmingen). Validation of antibody specificity for human and mouse endothelial antigens is shown in Supplementary Fig. 12.

Interphase FISH and FICTION on glioblastoma sections. Single- and dual-probe interphase FISH was performed on histological sections of glioblastoma, on cell nuclei extracted from paraffin-embedded sections of glioblastoma, on cells sorted from glioblastoma samples, and on cultured microvascular endothelial cells of glioblastoma as described[3]. Aneuploidy was definened as loss or gain of one or more chromosome FISH signals. Briefly, locus-specific probes for Cep10, Tel19q and LSI22 were used (Vysis). Standard FISH protocols for pretreatment, hybridization and analyses were followed according to the manufacturer's instructions. Histological 4-μm-thick paraffin sections were dewaxed with xylene and digested with proteinase K 1μg ml^{-1} in 0.002 M Tris buffered saline (TBS) for 20 min at room temperature (20 °C). Samples were then dehydrated in a graded ethanol series and subjected to FISH analysis. After specimen/probe denaturation at 73°C for 5 min, the probes (10 μl per slide) were applied to the slides and subsequently incubated overnight at 42 °C for Cep10 and at 37 °C for 10–16 h for LSI22/Tel19q. Post-hybridization procedure included subsequent washing in 50% formamide/2× SSC (30 min at 46 °C) and 2× SSC 0.1% NP40 (5 min at room temperature). Nuclei were counterstained with 4′,6-diamidino-2-phenylindole (DAPI; Vector Laboratories). The slides were studied with an Axioplan fluorescence microscope (Karl Zeiss) that was equipped with the appropriate filter sets (Vysis). Images were captured using a high-resolution black and white CCD microscope camera AxioCam MRm REV 2 (Karl Zeiss). The resulting images were reconstructed with green (FITC), orange and blue (DAPI) pseudocolour using AxioVision 4 multichannel fluorescence basic workstation (Karl Zeiss) according to the manufacturer's instruction. Glioblastoma sorted cells were fixed in a solution of methanol and acetic acid (3:1) for 10 min and then processed for FISH as described.

Laser capture microdissection of vessels from GSC-derived xenografts. We isolated the vascular structures of tumour xenografts using the Laser Capture Microdissection (LCM) System (PixCell IIe, Arcturus; distributed by Euroclone). LCM was performed on CD31-immunostained (M-20, Santa Cruz Biotechnology) paraffin sections (10-μm thick) of tumour xenografts. For each sample, laser power (50–70 mW) and laser duration (1–1.2 ms) were adjusted. The microdissected tissue was then transferred to an LCM cap and the cells were incubated in 100 ml digestion buffer (0.005% proteinase K in tris(hydroxymethyl)aminomethane (TRIS) 0.05M pH 7). Endothelial cell nuclei were isolated using the NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Scientific) following manufacturer recommendations. Successively, nuclei were washed with PBS and fixed in a solution of methanol/acetic acid (3:1). Eight millilitres of nuclei suspension were placed on a positive charged slide and were dried in a 65 °C oven for 30 min.

FISH on cell nuclei extracted from tumour xenografts. To distinguish human endothelial cells from mice cells, we performed FISH analyses using locus-specific probes for Cep10 (Vysis) and a Cy3-conjugate mouse pan-centromeric chromosome (Cambio). FISH protocols for the Cep10 probe were performed as previously described, whereas for the mouse pan-centromeric probe we followed the manufacturer's instructions. Briefly, after enzymatic digestion with 4mgml^{-1} pepsin in NaCl 0.9% pH 1.5 for 20 min at 37 °C, the nuclei were denatured in 70% formamide in 2× SSC for 2min at 70 °C, and were subsequently immersed in ice-cold 70% ethanol and dehydrated through a series of alcohol washes at 79%, 90% and 100%. The probe was denatured for 10 min at 85°C and immediately chilled on ice. After specimen/probe denaturation, probe was applied to the slide and subsequently incubated overnight at 37 °C. After washing, nuclei were then counterstained with DAPI (Vectashield mounting medium with DAPI; Vector Laboratories).

Isolation and culture of human glioblastoma microvascular endothelial cells. Glioblastoma tissue specimens were stored in medium M199 (Gibco) containing penicillin 100Uml^{-1} at 4 °C for less than 24 h before processing. After several washes with PBS/antibiotics, tissue was finely minced using surgical scissors and then incubated for 2–3 h at 37 °C in Dulbecco's medium (Gibco) containing 0.2% bovine serum albumin (BSA), liberase blendzyme 2–2.5 mg ml^{-1} (Roche Diagnostics). Cellular macroaggregates still present after enzymatic digestion were removed by filtration through a 10-μm pore-size filter (Dako), thus obtaining a monocellular suspension. The filtrate was then washed twice with PBS and centrifuged, the pellet resuspended in 1 ml cold PBS/0.1% BSA pH 7.4. Selection of endothelial cells was performed by using CD31 Miltenyi Microbead Kit (Miltenyi Biotech) according to manufacturer's instructions directly on cell suspensions after enzymatic digestion. Purified cell clusters as well as the negative counterparts were separately resuspended in endothelial basal growth medium (EBM Bullet kit; Biowhittaker Cambrex). Cells were plated onto 25-cm^2 culture dishes, previously coated with 1 μg cm^{-2} collagen type I and 1 μg cm^{-2} fibronectin (Sigma), and maintained at 37 °C in an atmosphere of 5% CO_2. After 10–12 h, plated cells were washed three times with cold PBS to favour detachment of nonendothelial cells. The medium was changed every 3 days. Once at confluence, cells were detached by trypsinization with 0.25% Trypsin/EDTA (Gibco) and reseeded on collagen/fibronectin-coated culture dishes at a split ratio of 1:3.A second magnetic selection was performed on plated endothelial cells after 7–10 cell divisions in order to increase the purity of the cultures.

Endothelial function assays. For *in vitro* three-dimensional tube formation assay, twelve microlitres of tail collagen were dropped onto glass coverslips and allowed to polymerize for 1 h at 37 °C. Cells were then seeded on top of the gels at 50,000 cells per well and allowed to incubate. Then endothelial basal medium was added and cells were cultured for 7 days.To quantify the tube formation, image-analysis techniques were used that measure the length of the tubes and the number of the connections. Data were photographically recorded daily. The average total length and mean total number of junctions for different endothelial cords were further analysed using the two-sided Mann–Whitney U test.For microinjections, a Zeiss microscopy with a manipulator was used. Fluorescein (Monico) was prediluted 1:1,000 into medium and injected into three-dimensional culture with a Hamilton, and observed with a Zeiss Axiovision device camera.

To determine the uptake of acetylated LDLs, cells were incubated with 10 mg ml^{-1} DiI-labelled (1,1′-dioetadeeyl-3,3,3′,3′-tetramethylindocarboeyanine perchlorate) acetylated LDL; Molecular Probes) at 37 °C for 4 h. The slides were analysed using a Nikon Eclipse TE300 inverted microscope equipped with a Zeiss Axiovision device camera.

Gene array. Total RNA was extracted from glioblastoma neurospheres, serum-differentiated glioblastoma neurospheres, glioblastoma neurospheres cultivated under endothelial condition and endothelial cells isolated from glioblastoma patients. Normal human umbilical vascular (HUVEC) or microvascular (HMVEC) endothelial cells were used as controls for endothelial gene expression patterns. RNA was labelled and hybridized to Affymetrix GeneChip1.0ST arrays following the manufacturer's instructions. Hybridization values were normalized by the robust multiarray averaging (RMA) method and hierarchical clustering, with average linkage method, was performed according to samples' gene expression profile. Full data were submitted to ArrayExpress under the accession number E-MEXP-2891.

Lentiviral infection. Selective targeting of the cells expressing endothelial phenotype was obtained by modifying the pRRLsin.Tie2p.TKiresGFP.spre lentiviral vector provided by L. Naldini[15,16]. Viral particle production and GSC infection were performed as previously described[23].

***In vivo* experiments.** Studies involving animals were approved by the Ethical Committee of the Catholic University School of Medicine in Rome. Nude athymic

and SCID mice (female, 4–5 weeks of age; Charles River) were used. For subcutaneous xenografts, cells were resuspended 1×10^6 in 0.1 ml of cold PBS, mixed with an equal volume of cold Matrigel (BD Bioscience), and injected into the flanks of nude mice. For intracranial xenografts, 2×10^5 cells in 5 μl of PBS were injected stereotactically onto the striatum. Mice were killed by 16–20 weeks after grafting to collect tumour xenografts. On ganciclovir treatment, no major toxicity was observed in vital organs.

doi:10.1038/nature09624

Glioblastoma stem-like cells give rise to tumour endothelium

Rong Wang[1,2,3], Kalyani Chadalavada[4], Jennifer Wilshire[5], Urszula Kowalik[1], Koos E. Hovinga[1,6], Adam Geber[1], Boris Fligelman[1], Margaret Leversha[4], Cameron Brennan[1,3,7] & Viviane Tabar[1,2,3]

Glioblastoma (GBM) is among the most aggressive of human cancers[1]. A key feature of GBMs is the extensive network of abnormal vasculature characterized by glomeruloid structures and endothelial hyperplasia[2]. Yet the mechanisms of angiogenesis and the origin of tumour endothelial cells remain poorly defined[3–5]. Here we demonstrate that a subpopulation of endothelial cells within glioblastomas harbour the same somatic mutations identified within tumour cells, such as amplification of *EGFR* and chromosome 7. We additionally demonstrate that the stem-cell-like CD133^{+} fraction includes a subset of vascular endothelial-cadherin (CD144)-expressing cells that show characteristics of endothelial progenitors capable of maturation into endothelial cells. Extensive *in vitro* and *in vivo* lineage analyses, including single cell clonal studies, further show that a subpopulation of the CD133^{+} stem-like cell fraction is multipotent and capable of differentiation along tumour and endothelial lineages, possibly via an intermediate CD133^{+}/CD144^{+} progenitor cell. The findings are supported by genetic studies of specific exons selected from The Cancer Genome Atlas[6], quantitative FISH and comparative genomic hybridization data that demonstrate identical genomic profiles in the CD133^{+} tumour cells, their endothelial progenitor derivatives and mature endothelium. Exposure to the clinical anti-angiogenesis agent bevacizumab[7] or to a γ-secretase inhibitor[8] as well as knockdown shRNA studies demonstrate that blocking VEGF or silencing *VEGFR2* inhibits the maturation of tumour endothelial progenitors into endothelium but not the differentiation of CD133^{+} cells into endothelial progenitors, whereas γ-secretase inhibition or *NOTCH1* silencing blocks the transition into endothelial progenitors. These data may provide new perspectives on the mechanisms of failure of anti-angiogenesis inhibitors currently in use. The lineage plasticity and capacity to generate tumour vasculature of the putative cancer stem cells within glioblastoma are novel findings that provide new insight into the biology of gliomas and the definition of cancer stemness, as well as the mechanisms of tumour neo-angiogenesis.

Blood vessels within GBM express a variety of markers, including CD31 and CD105 (also known as PECAM1 and ENG, respectively); CD105 is a proliferation-associated molecule expressed in angiogenic endothelium[9]. Quantitative analysis of 16 GBM specimens by fluorescence-activated cell sorting (FACS) and immunohistochemistry showed that more than 70% of CD105^{+} cells co-express CD31 (Fig. 1a, b), VEGFR2 (also known as KDR) and von Willebrand factor (also known as VWF), exhibit endothelial morphology, and labelling by DiI-AcLDL (1,1′-dioctadecyl-3,3,3′,3′-tetramethyl-indocarbocyanine perchlorate-labelled acetylated low density lipoproteins, ref. 10), suggesting an endothelial phenotype (Supplementary Fig. 1a). On average, ~5% of the total cell population expressed CD31 in normal brain and GBM specimens ($n = 7$), whereas CD105^{+} cells were essentially absent in normal brain (Supplementary Fig. 1b). CD105^{+} cells were also isolated by FACS from fresh GBM specimens and injected with a collagen matrix[11] into the flank of NOD-SCID mice. The resulting implants were composed of a network of vascular channels of human origin, expressed CD105 and CD31 and showed evidence of uptake of systemically injected lectin (Fig. 1c).

Whereas endothelial cells in GBMs are often classified as "hyperplastic"[2], the abnormal blood vessel architecture, the distinct gene expression profiles[12] and the selective emergence of abnormal vessels in GBMs versus lower grade gliomas[2] suggest a more complex ontogeny of GBM endothelium. We performed quantitative fluorescence *in situ* hybridization (FISH) analyses for *EGFR* and chromosome 7 (ref. 13) on CD105^{+} cells isolated by FACS and on sections of the corresponding GBM parent tumour (Fig. 1d, e and Supplementary Fig. 2). The proportion of CD105^{+} cells harbouring ≥3 copies of the *EGFR* amplicon or the centromeric portion of chromosome 7 was comparable to the proportion of tumour cells with the same aberrations (Supplementary Table 1a). We also performed quantitative PCR (qPCR) for three segments of the *EGFR* amplicon (exons 4, 9 and 11), known to be mutated at high frequency according to data from The Cancer Genome Atlas[6]. The data demonstrate a similar copy number in the CD105^{+} cells and the corresponding parent tumour (Supplementary Table 1b) and indicate that a proportion of tumour endothelial cells within GBM is in fact neoplastic.

CD133 is a cell surface glycoprotein used extensively as a marker of putative cancer stem cells (CSCs) but also expressed in haematopoietic stem cells[14–18]. Although the specific identity and definition of CSCs remains a matter of debate, we proposed that the CD133^{+} fraction may be related to the endothelial differentiation potential observed. Acutely dissociated cells from a series of 14 GBMs were fractionated into four groups: (1) CD144^{+}/CD133^{-}, (2) CD144^{+}/CD133^{+} (double positive, DP), (3) CD133^{+}/CD144^{-} and (4) CD133^{-}/CD144^{-} (double negative, DN) (Fig. 2a). All samples contained the four fractions, with the DN being the largest population (Supplementary Table 4). Quantitative PCR with reverse transcription (qRT–PCR) analysis for endothelial markers (Supplementary Fig. 3a) demonstrated marked enrichment of *VEGFR2* and the endothelial progenitor marker *CD34* in the CD144^{+}/CD133^{-} and in the DP populations. CD105 was consistently absent in the CD133^{+} and CD144^{+} fractions. To define lineage potential further, DP cells were cultured for 5 days in endothelial cell medium which resulted in the downregulation of CD144, the upregulation of CD105 and CD31 as well co-expression of VEGFR2 and CD34 and labelling with DiI-AcLDL (Fig. 2b and Supplementary Fig. 3b). When grown in three-dimensional (3D) gel cultures, the *in vitro* DP-derived endothelial cells form vascular networks reminiscent of normal endothelium, but also thickened channel walls and areas of confluence more suggestive of abnormal tumour vessels (Fig. 2c, d). The primary CD105^{+} cells also form glomeruloid-like structures in 3D gel, with high lectin uptake (Supplementary Fig. 1c). DP-derived CD105^{+} cells were sorted and injected subcutaneously in NOD/SCID mice, giving rise to

[1]Department of Neurosurgery, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [2]Center for Stem Cell Biology, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [3]Brain Tumor Center, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [4]Molecular Cytogenetics, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [5]Flow Cytometry Core, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [6]Neurosurgical Center Amsterdam, Academic Medical Center, Amsterdam 1105 AZ, The Netherlands. [7]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA.

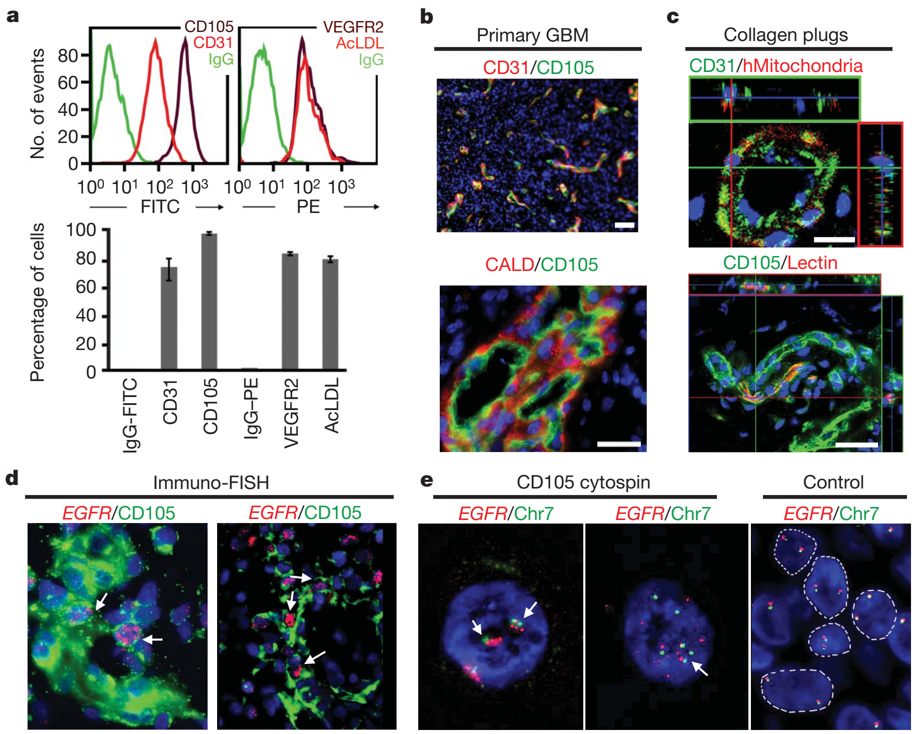


Figure 1 | CD105$^+$ endothelial cells in GBM harbour genomic aberrations. **a**, FACS analysis and quantification of GBM-derived CD105$^+$ cells shows co-expression of other endothelial cell markers (CD31, VEGFR2) and uptake of DiI-AcLDL($n = 3$). FITC, fluorescein isothiocyanate; PE, phycoerythrin **b**, CD105 immunostaining in GBMs delineates microvessels co-labelling with CD31 and glomeruloid vessels surrounded by caldesmon (CALD)-expressing pericytes. **c**, Functional neovessel formation by GBM-derived CD105$^+$ cells in the flank of NOD-SCID mice. Confocal immunofluorescence demonstrates co-localization of a human mitochondria marker with CD31 and uptake of lectin by the CD105$^+$ vessels in the implants. **d**, Immuno-FISH of CD105$^+$ vessels in GBM specimens (case 76, 78) shows multiple copies of the *EGFR* amplicon (arrows). **e**, FISH on CD105$^+$ cells sorted from GBMs confirms amplification of *EGFR* (red) and chromosome 7 centromere (Chr7, green) (arrows). Control nuclei, individually contoured, are from normal human fibroblasts. Scale bars, 50 µm. Error bars, s.d.

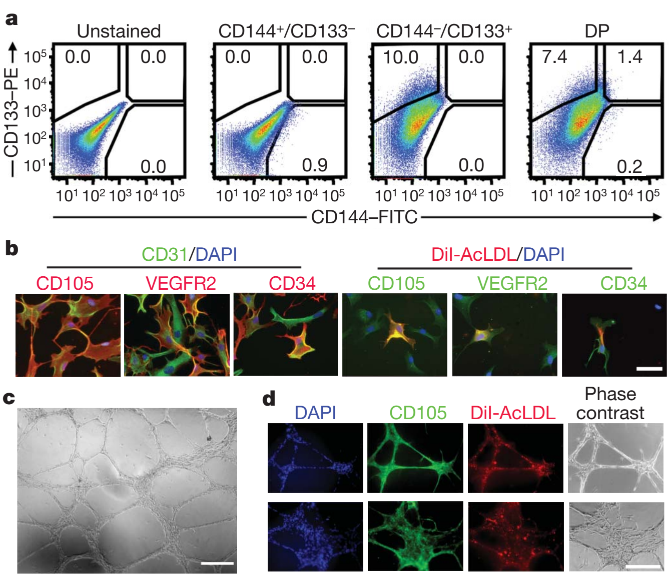


Figure 2 | GBM-derived CD133$^+$ cells include a fraction of endothelial progenitors a, Representative FACS analysis of a GBM specimen with fractionation into four cell subpopulations based on the expression of CD133 and vascular E-cadherin (CD144). **b**, Immunofluorescence analysis of DP (CD133$^+$/CD144$^+$) cells upon differentiation demonstrates co-expression of endothelial markers and DiI-AcLDL uptake. **c**, **d**, In Matrigel, DP cells will exhibit DiI-AcLDL uptake and form tubular networks comparable to those shown by normal endothelial cells, as well as areas of thickened walls where cells are more proliferative. Scale bars, 100 µm in **b** and **d**; 300 µm in **c**.

vascularized plugs identical to those obtained from primary CD105$^+$ cells (Supplementary Fig. 1d).

The CD144$^+$/CD133$^-$ cell fraction was often very small but showed a restricted differentiation and immunohistochemical profile (Supplementary Fig. 3c, d). When grown in Matrigel, the CD144$^+$/CD133$^-$ cells develop tubular, capillary-like structures[19] and no glomeruli (Supplementary Fig. 3e). CD144$^+$/CD133$^-$ cells do not express neural markers or form neurospheres, thus indicating a more restricted endothelial precursor cell identity (Supplementary Fig. 3f). Unsupervised clustering of transcriptome data was performed on several data sets including independent samples of the four sorted tumour subpopulations, as well as CD144$^+$ human embryonic stem-cell-derived endothelial precursors and bone-marrow-derived CD34$^+$ endothelial progenitors (Supplementary Fig. 3g). Taken together, these results indicate that GBMs comprise cell fractions capable of endothelial cell differentiation.

The identification of genomic aberrations in tumour endothelium and the presence of endothelial progenitors within the CD133$^+$ putative CSC fraction in GBMs, led us to postulate that DP cells may represent the neoplastic origin of tumour endothelium and could derive from the CD133$^+$ CSC fraction. CD133$^+$/CD144$^-$ cells were then labelled with *EFα-1::GFP* (elongation factor α1–green fluorescent protein) lentiviral vectors, triple sorted, and GFP$^+$/CD133$^+$/CD144$^-$ cells were co-cultured in the presence of tumour cells. On day 5, FACS analysis demonstrated the emergence of a GFP$^+$-DP population (Fig. 3a, b). When placed in collagen cultures, the GFP$^+$-DP cells had intracellular vacuoles suggestive of early lumen formation by endothelial tubes[20] (Fig. 3c), and differentiation into cells that express CD105 and CD31 and exhibit DiI-AcLDL uptake (Fig. 3d and Supplementary Fig. 4a). Importantly, co-culture with tumour cells is essential for the emergence of DP cells (Fig. 3a, b). Our data confirm that the DP endothelial

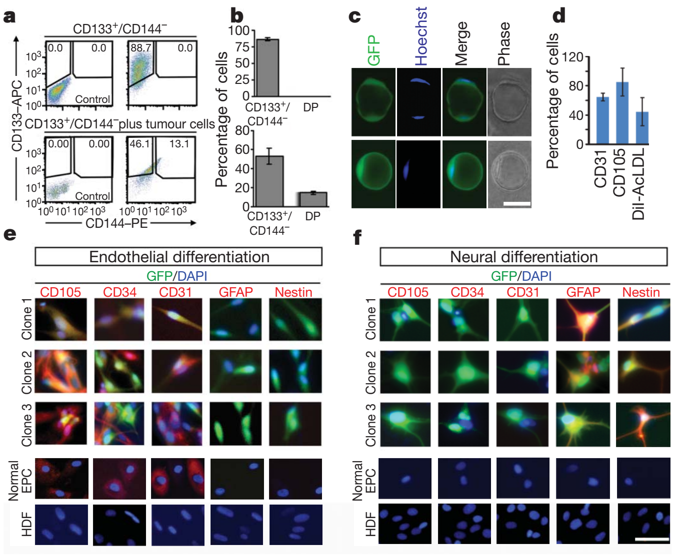


Figure 3 | CD133$^+$/CD144$^-$ cells are multipotential and give rise to endothelial cells via an endothelial progenitor intermediate. a, b, Co-cultures of CD133$^+$/CD144$^-$ cells with tumour cells give rise to endothelial progenitors that co-express CD133 and CD144 (DP) as shown and quantified by FACS analysis ($n = 3$). APC, allophycocyanin. **c**, GFP$^+$-derived DP cells form intracellular vacuolar structures in collagen gel, characteristic of endothelial cells. **d**, Immunohistochemistry of CD133$^+$/CD144$^-$-derived endothelial cells ($n = 3$). **e**, **f**, Single cell clonal analysis of GFP-labelled CD133$^+$/CD144$^-$ cells. GFP$^+$ clones derived from single cells are seeded under neural or endothelial conditions. Normal endothelial precursor cultures (EPC) and human dermal fibroblasts (HDF) were used as controls. Under endothelial conditions, all cells except HDF express endothelial but not neural markers. Under neural conditions, cells from the same GFP/CD133$^+$ clone are positive for GFAP and nestin but not endothelial markers, while controls are negative for all markers. Scale bar, 50 μm. Errors are s.d.

progenitors within GBM can arise from the CD133$^+$ cell population and are capable of differentiating into endothelial cells of tumour origin. Of note, the tumour cells used in these co-culture experiments originate from tumours with different genetic backgrounds and transcriptomal subclasses (Supplementary Tables 2 and 3).

Recent data support a close interaction[21] or a lineage relationship[22] between endothelial cells and neural stem cells. We next explored whether endothelial differentiation of CD133$^+$/CD144$^-$ can be further promoted by extrinsic signals. To this end, CD133$^+$/CD144$^-$ cells were isolated from GBM samples, stably transduced with *EFα-1::GFP* lentiviral vectors, sorted for GFP$^+$/CD133$^+$/CD144$^-$ and co-cultured with tumour-derived endothelial cells. GFP-expressing endothelial cells were identified at 7–10 days *in vitro* as demonstrated by co-labelling of GFP with CD105 and CD31, and also incorporation of DiI-AcLDL. Control experiments using GFP-labelled CD133$^-$ cells did not yield any endothelial cells (Supplementary Fig. 4b). The CD133$^+$/CD144$^-$ population formed neurospheres and readily differentiated along the three main CNS lineages (Supplementary Fig. 4c). Whereas these data are suggestive of the multipotent nature of the CD133$^+$/CD144$^-$ cells, they do not rule out the presence of heterogeneous populations within the CD133$^+$/CD144$^-$ fraction with distinct differentiation potentials. We thus performed single-cell clonal studies of CD133$^+$/CD144$^-$ cells as well as normal endothelial cells and fibroblasts as controls (Supplementary Fig. 4d). The data demonstrate both endothelial and neural differentiation potential within a single-cell derived clone confirming that CD133$^+$/CD144$^-$ cells are capable of generating tumour cells and tumour-derived endothelium (Fig. 3e, f). FISH for *EGFR* and chromosome 7 in the clones confirmed the presence of genomic amplifications identical to those exhibited by the parent tumour tissue (Supplementary Fig. 4e).

We next tested the fate of the various tumour cell fractions upon transplantation *in vivo*. CD133$^+$/CD144$^-$, DP, CD144$^+$/CD133$^-$ and DN cells were injected into the striatum of immunodeficient mice. All grafted animals developed tumours with the exception of those that received cells from the DN and CD144$^+$/CD133$^-$ fraction. DP and CD133$^+$/CD144$^-$ gave rise to large, highly infiltrative and hyperproliferative tumours showing strong expression of nestin (Fig. 4a). Whereas all xenograft tumours had a comparable volume and proliferation rate, the DP-derived tumours showed significantly increased levels of vascularization as demonstrated quantitatively (Supplementary Fig. 5a).

Some of the animals were grafted with stably GFP-marked CD133$^+$ cells allowing us to serially passage GFP-labelled CD133$^+$/CD144$^-$ cells from the primary xenograft in NOD-SCID mice. Secondary tumours formed at similar efficiency and showed comparable cell composition to the first passage cells. FACS analysis of GFP labelled xenograft cells demonstrates expression of endothelial markers, including CD105 and CD34 (Fig. 4b). After a second passage *in vivo*, tumours were sorted again for GFP$^+$/CD133$^+$/CD144$^-$ cells, which upon culture gave rise to GFP-labelled CD31$^+$ and CD105$^+$ cells, thus demonstrating maintenance of the multipotential phenotype (Fig. 4c). Immunohistochemical analysis, including confocal microscopy, demonstrated tumour blood vessels with typical morphology that express human markers. Tumour-bearing animals were also injected systemically with lectin, resulting in vessel-specific uptake and co-labelling with human markers (Fig. 4d, e and Supplementary Fig. 6). Thus, multipotency—including differentiation capacity along endothelial lineages—is maintained within the CD133$^+$/CD144$^-$ population *in vivo* and upon passaging. However, in the absence of clonal studies *in vivo*, true multipotency of tumour stem-like cells cannot be definitively confirmed.

A more comprehensive and quantitative analysis of genomic aberrations was conducted in order to verify the lineage relationship among the different tumour subpopulations. qPCR for the *EGFR* exons as described above[6] demonstrates the highest copy number within the CD133$^+$/CD144$^-$ population followed by the endothelial progenitors (DP) and the CD105$^+$ cells (Supplementary Fig. 5b). Interestingly, the CD31$^+$ cells and the CD144$^+$/CD133$^-$ progenitors showed lower levels of amplification, indicating that they may include a significant proportion of genotypically normal cells. We propose that these cells largely represent normal endothelium and circulating endothelial progenitors, respectively. This is compatible with the more restricted endothelial fate demonstrated by the CD144$^+$/CD133$^-$ cells as shown above (Supplementary Fig. 3c, e, f). Quantitative FISH studies for copy number of *EGFR* and chromosome 7 per cell were performed on CD133$^+$/CD144$^-$, DP and CD105$^+$ cells and revealed a substantial proportion of cells bearing the neoplastic aberrations in each population, ranging from 47.3% to 71.7% (Supplementary Fig. 5c). To address genomic alterations in tumour cells in a more unbiased manner we performed array comparative genomic hybridization (CGH) on the fractionated populations (Supplementary Fig. 7). The CGH data showed similar patterns of genomic aberrations in tumour cells as well as the endothelium and its progenitors, at variable amplitudes and across different regions, thus demonstrating a similar paradigm even in tumours that do not exhibit *EGFR* gain. We performed transcriptome analyses on a set of 18 tumours used in this study and found a random distribution of commonly described genotypes as well as representation of all TCGA-defined transcriptomal classes (Supplementary Table 3). Finally, we performed metaphase spreads on purified cell fractions of CD133$^+$/CD144$^-$, DP and CD105$^+$ following short-term culture. The majority of the cells had a highly abnormal but near-diploid karyotype, indicating that nuclear fusion is a very unlikely explanation for the lineage transition from cancer cell to endothelial progenitor or mature cell (Supplementary Fig. 5d). Vascular mimicry has been described in melanoma[23] and other tumours[24]; aneuploidy was also shown in renal cell cancer endothelium, but not matched to parent tumour cells[25].

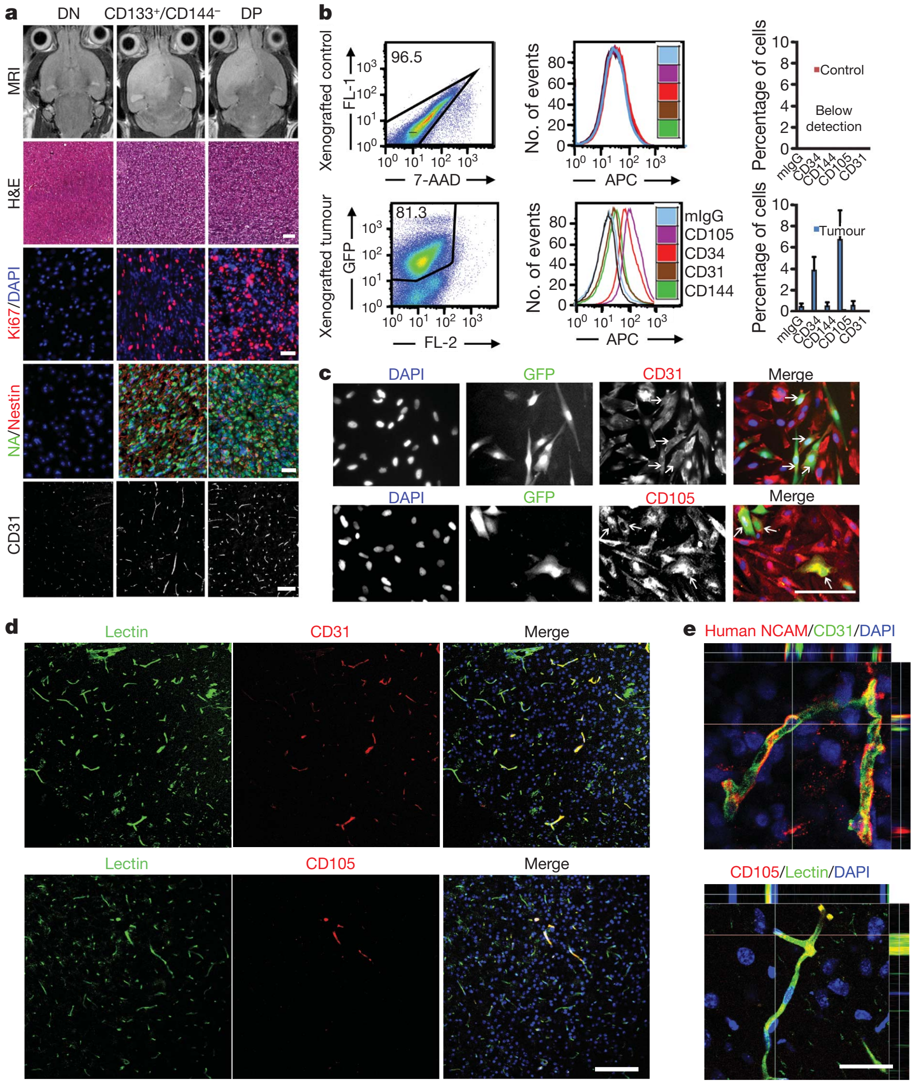


Figure 4 | Cancer stem-like cells and endothelial progenitors give rise to tumour and endothelial cells *in vivo*. **a**, Representative magnetic resonance imaging (MRI) images from mice that received injection of DN, CD133$^+$/CD144$^-$ or DP cells from primary GBM specimens. T2 sequences demonstrate infiltrative tumours except in the DN group. Tumours were hypercellular on haematoxylin and eosin (H&E), showed high proliferation rates (Ki67) and nestin expression. Immunostaining for human-specific CD31 demonstrates the presence of vessels of human origin within the tumours. NA, human nuclear antigen. **b**, FACS plots (left) and quantitative analysis (right) for endothelial marker expression in xenograft tumours (GFP$^+$/CD133$^+$/CD144$^-$ cells) and controls (DN). ($n = 3$, s.d.). 7-AAD, 7-aminoactinomycin. FL-1 and 2, fluorescent channels 1 and 2; mIgG, mouse immunoglobulin G. **c**, Xenograft derived GFP$^+$/CD133$^+$/CD144$^-$ cells express endothelial markers upon *in vitro* differentiation (arrows). **d**, Uptake of systemic lectin in tumour xenografts demonstrates blood vessels that co-label with human endothelial markers (CD31 and CD105). **e**, Confocal microscopy of xenograft microvasculature. Scale bars, 100 μm in **a**; 50 μm in **c**; 140 μm in **d**; 10 μm in **e**.

We investigated the impact of DAPT (N-[N-(3,5-difluorophenacetyl)-L-alanyl]-*S*-phenylglycine *t*-butyl ester), a γ-secretase inhibitor that effectively inhibits Notch signalling[8], and bevacizumab, a VEGFA-binding antibody[7] currently in clinical use, on the differentiation of CD133$^+$/CD144$^-$ to DP and then to CD105$^+$ cells. Exposure to bevacizumab did not have an impact on the ability of CD133$^+$/CD144$^-$ cells to differentiate into endothelial progenitors, yet it blocked further maturation from DP into CD105$^+$ endothelial cells. In contrast, γ-secretase inhibition resulted in significant suppression of the transition from CD133$^+$/CD144$^-$ to DP, but did not affect maturation to CD105$^+$ cells. To demonstrate the specific roles of the Notch and VEGF pathways, we performed knockdown studies targeting the *NOTCH1* and *VEGFR2* receptors. The gene silencing data further supported the results of the inhibitor studies (Supplementary Figs 8b and 9). Gene expression analysis shows significant upregulation of *NOTCH1/2* and *VEGFR1/2* in the CD133$^+$/CD144$^-$ and DP groups, respectively (Supplementary Fig. 8). These preliminary studies offer a novel perspective of the roles of the VEGF and Notch pathways in glioma biology, although the functional consequences of VEGF or Notch blockade remain to be determined.

Despite some promise, bevacizumab therapy is often interrupted by GBM progression characterized by a decrease in abnormal vascularity and significant invasive tumour behaviour[26]. Based on the paradigm presented here (Supplementary Fig. 9a), bevacizumab failure could be conceivably due to the disruption of the dynamic relationships between the tumour fractions.

In summary, our data demonstrate that a subpopulation of cells within GBM can give rise to endothelial cells via a bipotential progenitor intermediate, and that the CD133$^+$ cancer stem-cell-like fraction includes a population of endothelial progenitors. An in-depth understanding of the lineage relationship between tumour cells and endothelial progeny should provide new insights into CSC biology and tumour self-renewal. Given the strong correlation of tumour grade and neoplastic vasculature in human gliomas, agents that could block endothelial

transition of tumour cells may provide a novel therapeutic strategy for this currently intractable disease.

METHODS SUMMARY

All experiments were conducted on freshly obtained surgical specimens of glioblastoma tumour; a neuropathologist confirmed the diagnosis on frozen section before tissue acquisition. Tumours were newly diagnosed or recurrent. A total of 78 tumours were used in the study. Cell fractions were sorted using standard methods at our FACS facility; *in vitro* experiments were conducted on short-passage cultures (maximum of five passages) if needed. A total of 34 xenografts were obtained in immunodeficient mice following intrastriatal implantation of cell populations as indicated in the Methods. A lentiviral vector expressing GFP under a PGK promoter (gift from M. Sadelain) was used for cell labelling and sorting. Cytogenetic analyses were conducted using standard methods at the Cytogenetics Core facility at Sloan Kettering Cancer Center. Knockdown experiments were performed using lentiviral vectors expressing shRNA for *NOTCH1* or *VEGFR2* (Santa Cruz). All experiments were carried out in triplicates or greater. Data are expressed as mean ± s.d. P values were determined following two-tailed student's t-test. A P value of <0.05 was considered significant. Tissues were obtained after patients' written consent under a protocol approved by the institution's Institutional Review Board.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 May; accepted 1 November 2010.

Published online 21 November; corrected 9 December 2010 (see full-text HTML version for details).

1. Stupp, R. *et al.* Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* **352,** 987–996 (2005).
2. Kleihues, P. *et al.* The WHO classification of tumors of the nervous system. *J. Neuropathol. Exp. Neurol.* **61,** 215–225 (2002).
3. Kioi, M. *et al.* Inhibition of vasculogenesis, but not angiogenesis, prevents the recurrence of glioblastoma after irradiation in mice. *J. Clin. Invest.* **120,** 694–705 (2010).
4. Lyden, D. *et al.* Impaired recruitment of bone-marrow-derived endothelial and hematopoietic precursor cells blocks tumor angiogenesis and growth. *Nature Med.* **7,** 1194–1201 (2001).
5. Du, R. *et al.* HIF1alpha induces the recruitment of bone marrow-derived vascular modulatory cells to regulate tumor angiogenesis and invasion. *Cancer Cell* **13,** 206–220 (2008).
6. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455,** 1061–1068 (2008).
7. Kreisl, T. N. *et al.* Phase II trial of single-agent bevacizumab followed by bevacizumab plus irinotecan at tumor progression in recurrent glioblastoma. *J. Clin. Oncol.* **27,** 740–745 (2009).
8. Hovinga, K. E. *et al.* Inhibition of Notch Signaling in Glioblastoma Targets Cancer Stem Cells Via an Endothelial Cell Intermediate. *Stem Cells* **28,** 1019–1029 (2010).
9. Dallas, N. A. *et al.* Endoglin (CD105): a marker of tumor vasculature and potential target for therapy. *Clin. Cancer Res.* **14,** 1931–1937 (2008).
10. Voyta, J. C., Via, D. P., Butterfield, C. E. & Zetter, B. R. Identification and isolation of endothelial cells based on their increased uptake of acetylated-low density lipoprotein. *J. Cell Biol.* **99,** 2034–2040 (1984).
11. Laib, A. M. *et al.* Spheroid-based human endothelial cell microvessel formation *in vivo*. *Nature Protocols* **4,** 1202–1215 (2009).
12. Beaty, R. M. *et al.* *PLXDC1* (*TEM7*) is identified in a genome-wide expression screen of glioblastoma endothelium. *J. Neurooncol.* **81,** 241–248 (2007).
13. Verhaak, R. G. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*. *Cancer Cell* **17,** 98–110 (2010).
14. Singh, S. K. *et al.* Identification of human brain tumour initiating cells. *Nature* **432,** 396–401 (2004).
15. Richardson, G. D. *et al.* CD133, a novel marker for human prostatic epithelial stem cells. *J. Cell Sci.* **117,** 3539–3545 (2004).
16. Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc. Natl Acad. Sci. USA* **100,** 3983–3988 (2003).
17. Shmelkov, S. V., St, C. R., Lyden, D. & Rafii, S. AC133/CD133/Prominin-1. *Int. J. Biochem. Cell Biol.* **37,** 715–719 (2005).
18. Uchida, N. *et al.* Direct isolation of human central nervous system stem cells. *Proc. Natl Acad. Sci. USA* **97,** 14720–14725 (2000).
19. Koh, W., Mahan, R. D. & Davis, G. E. Cdc42- and Rac1-mediated endothelial lumen formation requires Pak2, Pak4 and Par3, and PKC-dependent signaling. *J. Cell Sci.* **121,** 989–1001 (2008).
20. Kamei, M. *et al.* Endothelial tubes assemble from intracellular vacuoles *in vivo*. *Nature* **442,** 453–456 (2006).
21. Shen, Q. *et al.* Endothelial cells stimulate self-renewal and expand neurogenesis of neural stem cells. *Science* **304,** 1338–1340 (2004).
22. Wurmser, A. E. *et al.* Cell fusion-independent differentiation of neural stem cells to the endothelial lineage. *Nature* **430,** 350–356 (2004).
23. Hendrix, M. J. *et al.* Transendothelial function of human metastatic melanoma cells: role of the microenvironment in cell-fate determination. *Cancer Res.* **62,** 665–668 (2002).
24. El Hallani, S. *et al.* A new alternative mechanism in glioblastoma vascularization: tubular vasculogenic mimicry. *Brain* **133,** 973–982 (2010).
25. Akino, T. *et al.* Cytogenetic abnormalities of tumor-associated endothelial cells in human malignant tumors. *Am. J. Pathol.* **175,** 2657–2667 (2009).
26. Iwamoto, F. M. *et al.* Patterns of relapse and prognosis after bevacizumab failure in recurrent glioblastoma. *Neurology* **73,** 1200–1206 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank J. Imai, H. Xu, G. Lee, M. Tomishima and L. Studer for critical reading of the manuscript, P. Gutin for assistance with tissue acquisition and discussions, B. Weksler for the brain endothelial cell line (hCMEC), S. Jhanwar for the clinical cytogenetics data and M. Sadelain and E. Papapetrou for the lentiviral vectors. Funding was provided in part through a grant from the New York State Stem Cell Science Fund (NYSTEM).

Author Contributions R.W. and V.T. conceived the project, analysed the data and wrote the manuscript. R.W. and remaining authors performed experiments and analysed data.

Author Information Microarray and CGH data are deposited in NCBI's Gene Expression Omnibus (GSE24244, GSE24446, GSE24452, GSE24557 and GSE24558). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to V.T. (tabarv@mskcc.org).

METHODS

Tissue processing. Surgical specimens were collected from the surgical suite at Memorial Sloan Kettering Cancer Center, following diagnostic confirmation by a neuropathologist. Tissues were obtained after patients' written consent under a protocol approved by the institution's Institutional Review Board. Tumours were cut mechanically first (McIlwain Tissue Chopper) then dissociated into single cells with Liberase Blendzyme 1 (Roche) as described previously[27].

Single cells were blocked with human FcR (1:20, Miltenyi Biotec) at 4 °C for 20 min before incubation with primary antibodies for 30 min. Cells from xenografts were further blocked with 2,4-G2 (1:100; Santa Cruz Biotechnology) before incubation with antibodies. Cells were incubated with primary antibodies, washed and reincubated with appropriate secondary antibodies and resuspended in FACS buffer[27] (containing 1× $Ca2^+/Mg2^+$-free HBSS (Invitrogen), 10 mM of HEPES, 0.156% of glucose and 0.5% of low endotoxin BSA fraction V, all from Sigma (Sigma-Aldrich), at a pH of 7.2) with 1 μg ml^{-1} 7-aminoactinomycin D (7-AAD, BD Pharmingen) before analysis. Mouse IgG1 or secondary antibody alone served as control for unspecific binding. Samples were analysed on a FACS Aria flow cytometer with CellQuest software (BD Biosciences) and data were analysed using FlowJo software (Tree star). A minimum of 10,000 events were counted and cell surface expression was analysed in 7-AAD-negative live cells. Antibodies used include: phycoerythrin- or allophycocyanin-conjugated anti-CD133 (1:20, Miltenyi Biotec); FITC-conjugated anti-CD144 (1:20, Abcam), anti-CD105 (1:20, BD Biosciences) and anti-CD31 (1:20, BD Biosciences); mouse anti-human CD31 (1: 40; BD Biosciences), mouse anti-human CD105 (1:40; Dako), mouse anti-human VEGFR2 (1:40; Abcam), mouse anti-human CD34 (1:20; Abcam), mouse anti-human CD144 (1:20; Abcam); mouse anti-human CD133 antibodies (AC141 and AC133 epitopes (1:20 each), Miltenyi). FITC-conjugated lectin and tetramethyl rhodamine isothiocyanate (TRITC)-conjugated lectins were purchased from Vector and Sigma separately.

DNA and RNA preparation. FACS-sorted cell populations from 21 glioblastoma patients were used to extract total RNA using an Absolutely RNA Nanoprep kit (Stratagene) or an RNeasy Kit (Qiagen). All RNA samples were pre-treated with DNase. Sorted cell populations from eight glioblastoma patients were used to isolate genomic DNA using the Picopure DNA extraction kit (Molecular Devices), followed by phenol (Invitrogen) extraction.

***In vivo* studies.** Adult female NOD/SCID or male NOD/SCID gamma (NSG) mouse (Jackson Laboratory) were anaesthetized with ketamine/xylazine (Hospira) and placed in a stereotaxic frame (Stoelting Company). Freshly sorted cells were injected into the right striatum immediately after sorting at the following coordinates (relative to bregma): AP = +0.5, ML = −2, and DV = −2.7. Animals received 10,000 cells each of $CD133^+/CD144^-$, $CD133^-/CD144^+$ or DP. NSG mice received 500 cells each of $CD133^+/CD144^-$ or DP. DN cells were used in three separate doses (10,000, 50,000 and 100,000 cells). Animals were killed upon exhibiting symptoms. Some animals received FITC-conjugated lectin by retro-orbital injection before killing. Total animals grafted $n = 40$.

The gel implantation assay was modified from ref. 11. Briefly, GBM- or DP-derived $CD105^+$ (10^6 or 2×10^6 per ml) were resuspended in Collagen IV (Chemicon). GBM- or DP-derived cell-gel suspension (500 μl) was injected subcutaneously below the xiphoid in four or three mice separately. Some animals received TRITC-conjugated lectin by tail vein injection before killing. After transplantation (21 days) the implants were retrieved, fixed overnight in 4% (v/v) buffered formalin at 4 °C, embedded in Optimal Cutting Temperature Compound (O.C.T. compound, Sakura Finetek) and sectioned on a freezing cryostat (Leica) for histological examination. Animals were housed and cared for in accordance with the National Institutes of Health (NIH) guidelines for animal welfare and all animal experiments were performed in accordance with protocols approved by our Institutional Animal Care and Use Committee (IACUC).

Animal imaging. *In vivo* magnetic resonance imaging was performed on a Bruker Biospec 4.7-Tesla 40-cm horizontal bore magnet. The system is equipped with a 200 mT m^{-1} gradient system. Examinations were conducted using a 72-mm birdcage resonator for excitation, and detection was achieved using a 3 cm surface coil. T2-weighted spin echo images were acquired consecutively using a rapid-acquisition relaxation enhanced sequence (RARE). Animals were anesthetized with 2% isoflurane in N_2/O_2 mixture.

Immunofluorescence. Primary antibodies were chicken anti-GFP (1:1,000; Chemicon), mouse anti-human CD31 (1:400; Abcam); mouse anti-human CD34 (1:400; Abcam), mouse anti-human CD105 (1:400; Dako); mouse anti-human vWF (1:100; Dako); mouse anti-human VEGFR2 (1:200; Abcam); mouse anti-human Ki67 (1:400; Dako); mouse anti-human NCAM (1:150; Santa cruz Biotechnology), mouse anti-human mitochondria (1:200; Chemicon), mouse anti-human nestin (1:400; Millipore), mouse anti-human nuclear antigen (1:500; Chemicon), rabbit anti-human GFAP (1:1,000; Chemicon), mouse anti-human Tuj1 (1:500; Covance), mouse anti-O4 (1:200; Chemicon), rabbit anti-human caldesmon (1:400; Novus Biology). The following secondary antibodies were used: Alexa Fluor 488-conjugated goat anti-chicken or mouse or rabbit (1:1,000), Alexa Fluor 555-conjugated goat anti-mouse or rabbit (1:1,000), Alexa Fluor 555-conjugated goat anti-mouse IgM (1:500), all from Molecular Probes (Invitrogen).

Cell culture and clonal assays. GFP labelling was obtained by incubation with a PGK-GFP lentiviral vector (gift from M. Sadelain). For sphere cultures, freshly sorted $CD133^+/CD144^-$, DP and $CD133^-/CD144^+$ cells were cultured under clonal conditions (1,000 cells per cm^2 or 5 cells per μl) in low-adherence plates (Corning) and maintained in serum free-Neurobasal medium supplemented with N2 (Invitrogen), 2 mM L-glutamine, 20 ng ml^{-1} recombinant human epidermal growth factor, and 10 ng ml^{-1} recombinant human fibroblast growth factor 2 (all from Invitrogen). Neurospheres were reseeded every 5 days after dissociation with Accutase (Innovative Cell Technologies). For neural differentiation, $CD133^+/CD144^-$ cells were cultured in laminin coated plates (BD Biosciences) using NeuroCult NS-A Differentiation Kit (human) (Stem Cell Technologies).

For endothelial progenitor cells, freshly sorted DP or $CD144^+/CD133^-$ cells were seeded on human fibronectin-coated plates (BD Biosciences) at a density of 10^5 ml^{-1} with endo-cult liquid medium Kit (Stem Cell Technologies) for propagation. DP, $CD144^+/CD133^-$ or $CD133^+/CD144^-$-derived DP cells were grown to 75% confluence and switched to M199 medium (Invitrogen) for quantification of endothelial differentiation as described previously[19]. GBM-derived $CD105^+$ cells were grown in M199 medium[19] for 2 days before FACS analysis. The functional assay for endothelial cells was performed by incubation of cells with 10 μg ml^{-1} of DiI-labelled acetylated low density lipoproteins (DiI-AcLDL) (Molecular Probes, Invitrogen) for 4 h.

For DP induction culture, GFP-labelled $CD133^+/CD144^-$ were co-cultured with tumour cells at a 20:1 ratio in N2 medium. The $CD133^+/CD144^-$-derived DP cells were sorted by FACS after 5 days for further characterization. A minimum of 100 cells were counted in triplicate assays. They were cultured in three-dimensional collagen gel[20]. For differentiation of $CD133^+/CD144^-$ to endothelial cells, tumour endothelial cells and GFP labelled $CD133^+/CD144^-$ cells or control cells were resuspended in endo-cult medium and grown on fibronectin coated plates for 7 days at a ratio of 100:1. Single cell clonal assays were performed by seeding freshly sorted single GFP-labelled cells on multi-well plates. Wells containing single green cells were identified and monitored until clone formation is established. Single-cell-derived clones were further sub-cloned and propagated twice, dissociated and seeded under neural and endothelial differentiation conditions as described above. Human umbilical cord-derived $CD133^+$ endothelial precursor cells (Biochain) or human dermal fibroblasts (Cell Applications) were maintained as per manufacturer instructions and used as control in clonal analysis.

Inhibitor studies. For drug treatment assays, cells were cultured in DP induction medium or endothelial differentiation medium containing 5 μM of the γ-secretase inhibitor DAPT (N-[N-(3,5-difluorophenacetyl)-L-alanyl]-*S*-phenylglycine *t*-butyl ester, Sigma-Aldrich) or 1 μg μl^{-1} of bevacizumab (Genentech). Treated cells were analysed by FACS analysis after 48 h incubation. VEGF was measured in the culture medium with a human VEGF ELISA Kit (Invitrogen) following the manufacturer directions.

Knockdown studies. GBM-derived fresh DP and GFP-CD133sp cells were infected with shRNA virus targeting *VEGFR2* or *NOTCH1* or a control virus (all from Santa Cruz Biotechnology). NOTCH1 shRNA lentiviral vector mix contains three target-specific constructs: CACCAGTTTGAATGGTCAATTCAAGAGATTGACCATTCAAACTGGTGTTTTT; CCCATGGTACCAATCATGATTCAAGAGATCATGATTGGTACCATGGGTTTTT; CCATGGTACCAATCATGAATTCAAGAGATTCATGATTGGTACCATGGTTTTT. VEGFR2 shRNA lentiviral vector mix contains three target-specific constructs: ACTGTGGTGATTCCATGTCTTCAAGAGAGACATGGAATCACCACAGTTTTTT; ACTTGTAAACCGAGACCTATTCAAGAGATAGGTCTCGGTTTACAAGTTTTTT; CACCTGTTTGCAAGAACTTTTCAAGAGAAAGTTCTTGCAAACAGGTGTTTTT. The infected cells were selected with 2–4 μg ml^{-1} puromycin (Santa Cruz Biotechnology) and used for FACS analysis and/or collected for RT–PCR as described above after 5 days in selection.

***In vitro* angiogenesis assay.** Intracellular vacuole formation was evaluated by culturing $CD133^+/CD144^-$-derived DP cells in three-dimensional collagen gel as described in ref. 20. Tubular network formation was assessed by culture in growth factor reduced Matrigel assay Kit (BD Biosciences) following the protocol from ref. 19.

Cytogenetic analyses and genomic PCR. Fluorescence *in situ* hybridization was performed using BAC clone RP11-339F13 and PAC clone RP5-1091E12 spanning the *EGFR* locus in 7p11, both labelled with Red-dUTP, together with a chromosome 7 centromere repeat DNA probe labelled with Green-dUTP targeted at the centromeric region of chromosome 7 (7p11.1–7q11.1 D7Z1 alpha satellite region). FISH was performed on sorted cells post cytospin on glass slides. A minimum of 100 cells in interphase were analysed. Human dermal fibroblasts

(HDF) served as normal control. The false positive rates for FISH probes was determined as 1% (s.d. = 1.3) and the cut-off level for the diagnosis of amplification was set at 5% (>3s.d.) ($n = 3$, total counted 2,000 control cells).

FISH on tumour sections, as reported in Supplementary Tables 1b and 2, was performed independently by the Clinical Cytogenetics Facility at Memorial Sloan Kettering Cancer Center, as part of a now routine molecular diagnostic test. The probe used is the 7p12 LSI EGFR and the 7p11.1–7q11.1 CEP (D7Z1 alpha satellite) dual colour probe, purchased from Abbott Molecular.

Fluorescence immunophenotyping and interphase cytogenetics, a technique combining immunohistochemistry for CD105 and FISH for *EGFR*, was carried out on 10-μm thick tissue sections. Normal human brain cerebral-cortex sections (Biochain) were used as controls.

In a copy number quantification reaction by real-time PCR, *EGFR* primers were designed based on published data[6,28,29,30]. Genomic DNA (10 ng) from sorted cells or normal human brain was used as template to examine the copy number of exons 4, 9, 11 in the *EGFR* gene; *GAPDH* was used as reference gene. Each replicate was normalized to *GAPDH* to obtain a ΔC_t, and then an average ΔC_t value for each sample (from the three replicates) was calculated. All samples were then normalized to the calibrator sample (normal human brain) to determine $\Delta\Delta C_t$. Relative quantity (RQ) is $2^{\Delta\Delta Ct}$, and copy number is $2 \times$ RQ. The EGFR copy number in each population was defined by the average of copy number from three exons. Error bars indicate the range of the data from the three exons in each of the three samples.

Karyotype analysis was performed on metaphase spread of FACS-purified cell subpopulations that were in culture for 3 days. The cultures were treated with Colcemid (0.1 μg ml^{-1}) for 1.5 h before *in situ* metaphase preparation according to standard cytogenetics procedures.

CGH studies. Comparative genomic hybridization (CGH) assay was performed by hybridizing genomic DNA from sorted cells with 44K human genome CGH arrays, and frozen banked whole tumour on 244K and 1M human genome CGH arrays (all commercial arrays from Agilent). DNA from sorted cells was prepared as described above. DNA extraction, purification, labelling and hybridization were performed at Sloan Kettering Cancer Center's Genomics Core Facility according to the manufacturer's instructions. Log_2 ratios were normalized by Lowess against probe intensity and mean %GC of the genomic region mapped to by the probe. Segmentation of normalized $\log_2$ ratios was by Circular Binary Segmentation (CBS, R package DNAcopy).

A separate method was used to investigate whether an amplicon identified by CBS in one cell fraction might be present in a minor subpopulation in other cell fraction at a level not detected by CBS. A region of interest (ROI) is defined by the boundaries of the amplicon detected by CBS. Then this region is investigated in the CGH profiles of the other cell fractions as follows: the $\log_2$ ratios of the N probes under the ROI (within amplicon boundaries) are compared to $\log_2$ ratios of all the other probes in the entire chromosome by Student's t-test (one-tailed). The observed t-score is then compared to the distribution of t-scores obtained by equivalently testing all other sets of N neighbouring probes in the chromosome. The ROI is considered to be significantly gained if the observed t-score is seen or exceeded in less than 0.1% of all other chromosomal regions.

Expression microarray studies of whole tumours. Gene expression profiling was performed for a subset of 16 tumours for which additional frozen material was available using exon expression arrays (Human Exon 1.0, Affymetrix). RNA was extracted, labelled and hybridized at Sloan Kettering Cancer Center's Genomics Core Facility according to the manufacturer's instructions. Data was normalized in a cohort of 80 gliomas using Aroma.affymetrix (R package aroma.affymetrix). Expression was derived for RefSeq transcripts, and multiple transcripts for the same gene were distilled to a single gene expression value by median. Transcriptomal class assignment was based on the nearest centroid of the four transcriptomal classes reported in ref. 13, using the subset of 840 signature genes described by this study (Supplementary Table 6; http://tcga-data.nci.nih.gov/docs/publications/gbm_exp/). Distances to centroids were defined using Pearson correlation and class assignments made by the largest correlation value. If the largest correlation was <0.2, the sample was labelled 'indeterminate'. Correlations and class assignments are given in Supplementary Table 6.

Microscopic analysis. Sections were examined with confocal laser scanning microscopy (Leica Microsystems; Carl Zeiss MicroImaging). The data was analysed with Velocity or LSM5 (Carl Zeiss MicroImaging) software.

Tumour microvessel density (MVD) was assessed by quantification of the numbers of CD31$^+$ tumour vessels in pixels using MetaMorph (Molecular Devices) image analysis software using unbiased sampling.

Gene expression analysis and quantitative real-time PCR for sorted cell populations. Total RNA of four subpopulations from two specimens were hybridized with human U133-plus2 array at Sloan Kettering Cancer Center's Genomics Core Facility and according to the manufacturer's instructions. Reference databases, including one set of CD34$^+$ human haematopoietic progenitor cells (GSM476781) and two independent sets of human embryonic stem cell-derived endothelial progenitors (GSM492830 and GSM492828) were downloaded from Gene Expression Omnibus database. The array data are analysed by Partek software. The data from 11 samples were normalized by RMA algorithms and the tumour samples then assigned in four groups based on the expression of membrane markers CD133 and CD144. The gene list is created by ANOVA with unadjusted P value less than 0.05 and then used as input for unsupervised hierarchical clustering by using Euclidian similarity metric.

For RT–PCR, total RNA (100–300 ng) was reverse-transcribed using random-primer and superscript III (Invitrogen) according to the manufacturer's instructions. Quantitative real-time PCR was performed with an Applied Biosystems Prism 7900HT sequence Detection System using SYBR Green PCR Master Mix (Applied Biosystems).

Primers: *CD34* (F: TCTGATCTCCATGGCTTCCT; R: ACTGAGGCAACAG CTCAACC), *CD144* (F: TCGTCATGGACCGAGGTT; R: TCTACAATCCCTT GCAGTGTGA), *VEGFR2* (F: GCAGGGGACAGAGGGACTTG; R: GAGGCC ATCGCTGCACTCA), *CD31* (F: TTCCTGACAGTGTCTTGAGTGG; R: GCT AGGCGTGGTTCTCATCT), *CD133* (F: TCTGGGTCTACAAGGACTTTCC; R: GCCCGCCTGAGTCACTAC), *ACTIN* (F: GCCCGCCTGAGTCACTAC; R: GGAATCCTTCTGACCCATGC), *VEGFR1* (F: TCTCACATCGACAAACCA ATACA; R: GGTAGCAGTACAATTGAGGACAAGA), *VEGF* (F: CTACCTCC ACCATGCCAAGT; R: CCACTTCGTGATGATTCTGC).

Human angiogenesis PCR arrays (SABiosciences) were used to examine the expression profiles of angiogenic genes in sorted cell populations. Heat Map construction and analysis of qPCR data was conducted according to ref. 31.

27. Panchision, D. M. *et al.* Optimized flow cytometric analysis of central nervous system tissue reveals novel functional relationships among cells expressing CD133, CD15, and CD24. *Stem Cells* **25,** 1560–1570 (2007).
28. Franco-Hernandez, C. *et al.* Gene dosage and mutational analyses of EGFR in oligodendrogliomas. *Int. J. Oncol.* **30,** 209–215 (2007).
29. Mellinghoff, I. K. *et al.* Molecular Determinants of the response of glioblastomas to EGFR kinase inhibitors. *N. Engl. J. Med.* **353,** 2012–2024 (2005).
30. Brennan, C. *et al.* Glioblastoma subclasses can be defined by activity among signal transduction pathways and associated genomic alterations. *PLoS ONE* **4,** e7752 (2009).
31. Guo, P. *et al.* Dual nature of the adaptive immune system in lampreys. *Nature* **459,** 796–801 (2009).

LETTER

doi:10.1038/nature09583

Calcium-dependent phospholipid scrambling by TMEM16F

Jun Suzuki[1,2], Masato Umeda[3], Peter J. Sims[4] & Shigekazu Nagata[1,2]

In all animal cells, phospholipids are asymmetrically distributed between the outer and inner leaflets of the plasma membrane[1]. This asymmetrical phospholipid distribution is disrupted in various biological systems. For example, when blood platelets are activated, they expose phosphatidylserine (PtdSer) to trigger the clotting system[2,3]. The PtdSer exposure is believed to be mediated by Ca^{2+}-dependent phospholipid scramblases that transport phospholipids bidirectionally[1,4], but its molecular mechanism is still unknown. Here we show that TMEM16F (transmembrane protein 16F) is an essential component for the Ca^{2+}-dependent exposure of PtdSer on the cell surface. When a mouse B-cell line, Ba/F3, was treated with a Ca^{2+} ionophore under low-Ca^{2+} conditions, it reversibly exposed PtdSer. Using this property, we established a Ba/F3 subline that strongly exposed PtdSer by repetitive fluorescence-activated cell sorting. A complementary DNA library was constructed from the subline, and a cDNA that caused Ba/F3 to expose PtdSer spontaneously was identified by expression cloning. The cDNA encoded a constitutively active mutant of TMEM16F, a protein with eight transmembrane segments[5]. Wild-type TMEM16F was localized on the plasma membrane and conferred Ca^{2+}-dependent scrambling of phospholipids. A patient with Scott syndrome[6,7], which results from a defect in phospholipid scrambling activity[8,9], was found to carry a mutation at a splice-acceptor site of the gene encoding TMEM16F, causing the premature termination of the protein.

When mouse Ba/F3 cells were treated with 1.0 μM A23187 for 15 min in the presence of 0.5 mM $CaCl_2$, the cells underwent necrosis or became propidium iodide (PI)-positive. However, when the same treatment was performed in Ca^{2+}-free conditions, most of the cells exposed PtdSer and the PI-positive population was low (Fig. 1a). Chelating intracellular Ca^{2+} with bis-(*o*-aminophenoxy)ethane-*N*,*N*,*N*′,*N*′-tetra-acetic acid acetoxymethyl ester (BAPTA-AM) blocked the PtdSer exposure (Fig. 1b), indicating that the process required the mobilization of intracellular calcium. This PtdSer exposure was reversible: treatment of the PtdSer-exposing cells with BAPTA-AM at 37 °C for 5 min (Fig. 1c) or culturing them in Ca^{2+}-free medium at 37 °C for 12 h (data not shown) eliminated the PtdSer from the cell surface. These results suggest that under low-Ca^{2+} conditions, A23187 mobilized the intracellular Ca^{2+}, which activated a phospholipid scramblase to expose PtdSer. When the intracellular Ca^{2+} concentration was lowered, the phospholipid scramblase lost activity, and flippases returned the PtdSer to the inner leaflet.

To characterize the PtdSer-exposure process, we used its reversible nature under low-Ca^{2+} conditions to establish a cell line that over-exposed PtdSer. Ba/F3 cells were treated with 1.0 μM A23187 in the absence of calcium, and subjected to fluorescence-activated cell sorting (FACS) based on PtdSer exposure. A population (0.5–5%) that showed intense staining with Annexin V was collected, cultured for 15 h in Ca^{2+}-free medium, returned to normal medium, and subjected to the next sorting. After this cycle of sorting and expansion had been repeated 12 times, the cells (Ba/F3-PS12) showed roughly 100-fold higher staining with Annexin V than the original Ba/F3 cells (Ba/F3-PS0) on treatment with 125 nM A23187 (Fig. 1d). The sorting and expansion were repeated another seven times, and the resulting cell line (Ba/F3-PS19) was used for further studies.

There were two possible causes of the strong PtdSer exposure in Ba/F3-PS19 cells. One was the overexpression or overactivation of phospholipid scramblase, and the other was the inactivation of flippase[10] that transports PtdSer from the outer to the inner leaflet of the plasma membrane. To examine which possibility was correct, DsRed-expressing Ba/F3-PS19 cells were fused with green fluorescent protein (GFP)-labelled parental Ba/F3 (Ba/F3-PS0) cells. The PtdSer-exposure response of the hybrid cells to 1.0 μM A23187 was similar to, or slightly weaker than, that of Ba/F3-PS19 cells (Fig. 1e), suggesting that the phenotype of Ba/F3-PS19 cells was dominant to that of Ba/F3-PS0 cells, and that the phospholipid scramblase was overactivated in Ba/F3-PS19 cells. To identify the gene responsible for the enhanced phospholipid scramblase activity, a cDNA library (9.3×10^5 clones) was prepared from Ba/F3-PS19 cells, and introduced into the parental Ba/F3 cell line. The stably transformed cells were treated with 125 nM A23187, and a population that stained strongly with Annexin V was sorted (Fig. 1f). At the third cycle of sorting and expansion (Library-Derived (LD)-PS3), about 35% of the cells exposed PtdSer without A23187 treatment, and this cell population (LD-PS4) was characterized.

LD-PS4 cells carried two or three different cDNAs, but the *Tmem16f* cDNA (GenBank accession number NM_175344) was present in two independent experiments, suggesting that TMEM16F caused the PtdSer exposure. The two *Tmem16f* cDNAs identified in the different experiments contained an A-to-G mutation at nucleotide 1226, which caused an aspartic residue to be replaced by glycine at codon 409 (Fig. 2a). TMEM16A, another member of the TMEM16 family, was recently shown to be a Ca^{2+}-dependent Cl^- channel[11–13]. However, the Cl^--channel activity of TMEM16F was lower than that of TMEM16A[14]. To examine the function of TMEM16F, the wild-type and mutant (D409G) forms of TMEM16F were tagged with Flag or monomeric red fluorescent protein (mRFP) at the carboxy terminus, and expressed in Ba/F3 or human 293T cells. Western blotting of the cell lysates with anti-Flag showed broad bands at 125 and 500 kDa on SDS–PAGE (Fig. 2b), suggesting that mouse TMEM16F (calculated molecular mass 106 kDa) is glycosylated and/or aggregated. Observation of the 293T cells expressing TMEM16F–mRFP indicated that TMEM16F is located at the plasma membrane (Fig. 2c).

Annexin V was able to bind to the Ba/F3 cells expressing the D409G mutant, but not the wild-type, TMEM16F (Fig. 2d), suggesting that the mutant TMEM16F-expressing cells constitutively expose PtdSer. This was confirmed by binding of MFG-E8, which specifically binds to PtdSer[15,16] (Supplementary Fig. 1). Chelating the intracellular Ca^{2+} with BAPTA-AM decreased the exposed PtdSer level in the mutant TMEM16F-expressing cells (Fig. 2d). When cells expressing wild-type TMEM16F were treated with A23187, PtdSer was exposed without a lag time, reaching saturation more quickly than the vector-transformed

[1]Department of Medical Chemistry, Graduate School of Medicine, Kyoto University, Yoshida, Sakyo-ku, Kyoto 606-8501, Japan. [2]Core Research for Evolutional Science and Technology, Japan Science and Technology Corporation, Kyoto 606-8501, Japan. [3]Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, Katsura, Nishikyo-ku, Kyoto 615-8510, Japan. [4]Department of Pathology and Laboratory Medicine, University of Rochester Medical Center, 601 Elmwood Avenue, Rochester, New York 14642, USA.

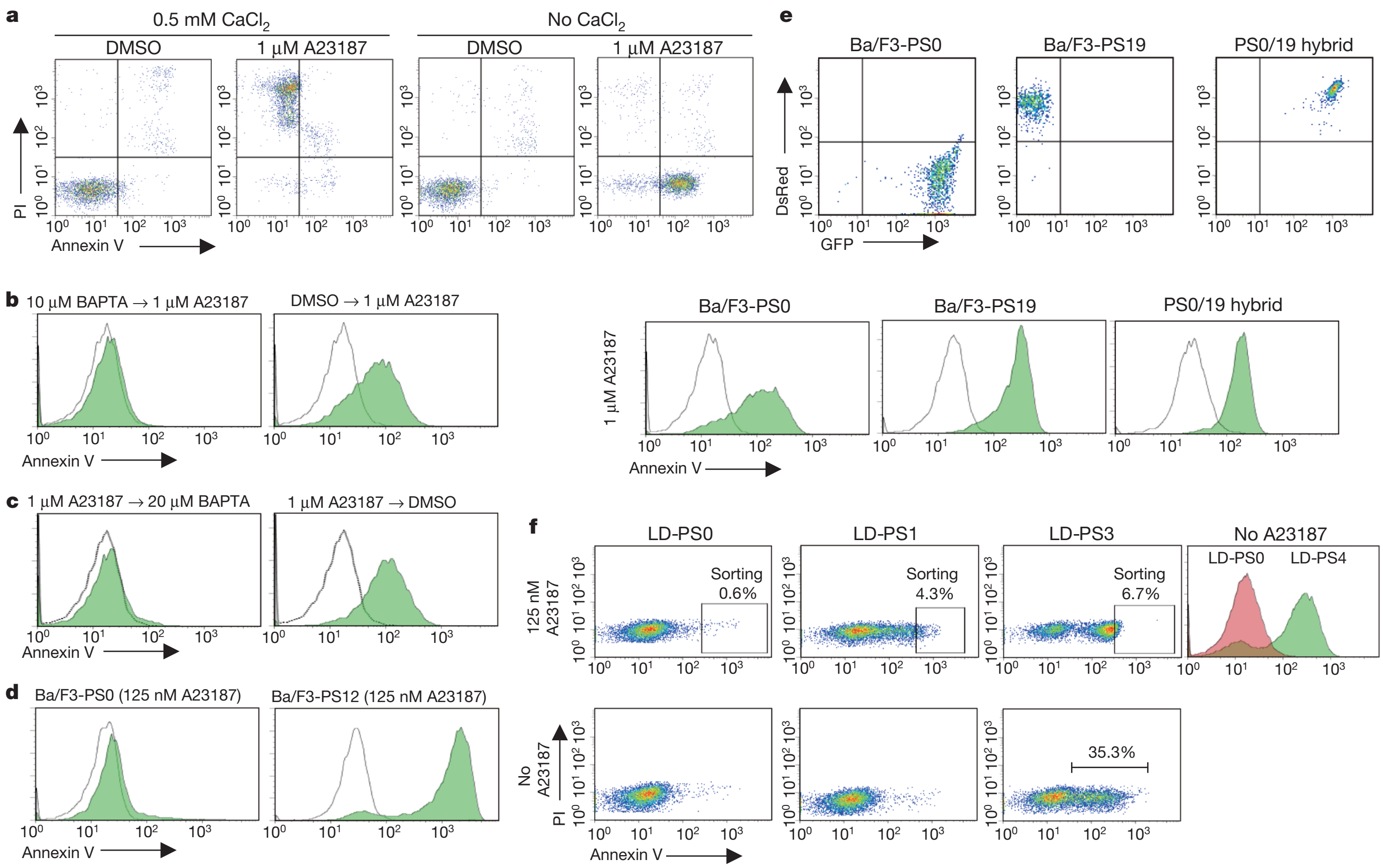


Figure 1 | Molecular cloning of TMEM16F. **a**, Ba/F3 cells were treated with A23187 with or without $CaCl_2$, and stained with Annexin V and PI. DMSO, dimethylsulphoxide. **b**, Ba/F3 cells were incubated with BAPTA-AM and treated with A23187. An Annexin V profile in a PI-negative population is shown. Open curve, profile of resting cells. **c**, Ba/F3 cells were treated with A23187 and then with BAPTA-AM for 5 min, and stained with Annexin V. **d**, Ba/F3 cells and cells after sorting for 12 cycles (PS12) were treated with A23187 and stained with Annexin V. **e**, GFP and DsRed profiles of PS0, PS19 and PS0/19 hybrid cells are shown. Bottom: the same cells were treated with A23187 and stained with Annexin V. **f**, Ba/F3 cells transformed with PS19 cDNA library were treated with A23187, stained with Annexin V and sorted (LD-PS0). Annexin V and PI profiles of cells after first (LD-PS1) and third (LD-PS3) sorting are shown. Right: Annexin V profile of original cells (LD-PS0) and after fourth sorting (LD-PS4) without A23187.

Ba/F3 cells (Fig. 2e). The intracellular Ca^{2+} concentration and the kinetics of the Ca^{2+} influx after treatment with A23187 was similar among the vector-transformed cells and those expressing wild-type and D409G mutant TMEM16F (Supplementary Fig. 2). These results indicated that TMEM16F mediates a Ca^{2+}-dependent scramblase activity for PtdSer, and that its D409G mutant is sensitized to respond to the normal intracellular concentration of Ca^{2+} to expose PtdSer.

Phospholipid scramblase mediates the bidirectional transfer between plasma membrane leaflets of all phospholipids. Cells expressing the D409G mutant TMEM16F were stained with Ro09-0198 (Supplementary Fig. 3a), a tetracyclic polypeptide that specifically binds phosphatidylethanolamine (PtdEtn)[17], indicating that they constitutively exposed PtdEtn, a phospholipid that, like PtdSer, is normally sequestered to the inner leaflet. Treatment of Ba/F3 cells with A23187 caused exposure of PtdEtn. This process was accelerated by overexpressing wild-type TMEM16F (Supplementary Fig. 3b). When 1-oleoyl-2-{6-[(7-nitro-2-1,3-benzoxadiazol-4-yl)amino]hexanoyl}-*sn*-glycero-3-phosphocholine (NBD-PtdCho) was added to the culture, it was quickly internalized by the D409G-mutant-expressing cells (Fig. 2f): of the cell-associated NBD-PtdCho, more than 40% became resistant to extraction with BSA within 6 min. When the cells expressing wild-type TMEM16F were treated with A23187, they incorporated NBD-PtdCho faster than the parental cells, and about 40% of the cell-associated NBD-PtdCho was inside the cells within 4 min (Fig. 2g). Similar results—that is, constitutive internalization by cells expressing the mutant TMEM16F, and enhanced A23817-induced incorporation by cells expressing wild-type TMEM16F—were obtained with *N*-{6-[(7-nitro-2-1,3-benzoxadiazol-4-yl)amino]hexanoyl}-sphingosine-1-phosphocholine (NBD-SM) (Supplementary Fig. 4). The internalized NBD-PtdCho and NBD-SM were intact (Supplementary Fig. 5). Dynasore, which inhibits dynamin-mediated endocytosis[18] inhibited the internalization of these phospholipids only slightly or not at all (Supplementary Fig. 6), suggesting that the contribution of endocytosis to TMEM16F-mediated phospholipid internalization may not be great.

Expression of endogenous TMEM16F in Ba/F3 cells was then knocked down by expressing *Tmem16f* short hairpin RNA (shRNA). As shown in Fig. 3a and Supplementary Fig. 7, the expression level of *Tmem16f* messenger RNA in five transformants was decreased to 20–35% of that in the cells expressing the control shRNA. The rate of A23187-induced exposure of PtdSer and PtdEtn was decreased in these transformants (Fig. 3b, c). Similarly, the uptake of NBD-PtdCho and NBD-SM was slower in *Tmem16f* shRNA-transformed cells (Fig. 3d, e).

Platelets and other blood cells from patients with Scott syndrome show a defect in their ability to expose PtdSer in response to a Ca^{2+} ionophore[7,19]. B-cell lines have been established from a patient with Scott syndrome and from the patient's parents[20]. In agreement with previous reports[8,20], the patient-derived cells did not expose PtdSer in response to a Ca^{2+} ionophore (Fig. 4a). In contrast, A23187 elicited PtdSer exposure in cell lines derived from the patient's parents at the same levels as in cell lines from healthy volunteers. An RT–PCR analysis of the *TMEM16F* mRNA (GenBank accession number NM_001025356) showed that the 5′ part (1,320 base pairs (bp)), corresponding to exons

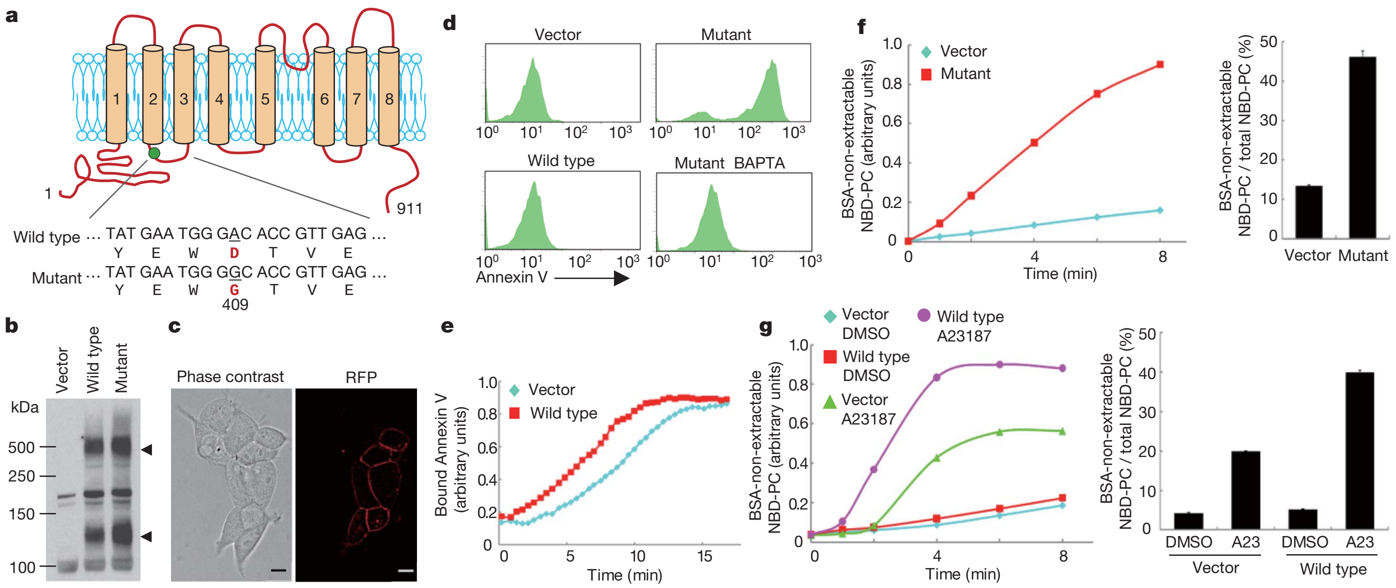


Figure 2 | Phospholipid scrambling in TMEM16F-expressing cells. **a**, Schematic representation of mouse TMEM16F and D409G mutant. **b**, Western blotting of Ba/F3 cells expressing Flag-tagged wild-type and mutant TMEM16F with anti-Flag. Arrowheads, monomer and multimer of TMEM16F. **c**, 293T cells expressing TMEM16F–mRFP were observed under a fluorescent microscope. Scale bars, 10 µm. **d**, Vector-transformed Ba/F3 cells, or cells expressing wild-type or mutant TMEM16F, were stained with Annexin V with or without pretreatment with BAPTA-AM. **e**, Vector-transformed or wild-type TMEM16F-Ba/F3 cells were preincubated with Annexin V. After addition of A23187, the fluorescence was monitored. The *y* axis shows fluorescence intensity on FACS. **f**, Vector-transformed or mutant TMEM16F-expressing Ba/F3 cells were incubated for 8 min at room temperature (26–27 °C) with 0.5 µM NBD-PtdCho in Hanks balanced salt solution containing Ca^{2+}. After dilution with fatty-acid-free BSA buffer, the fluorescence intensity was determined by FACS. **g**, Vector-transformed or wild-type TMEM16F-expressing Ba/F3 cells were preincubated at 4 °C with 0.1 µM NBD-PtdCho. A23187 (A23) was added and incubated for 8 min at room temperature, and internalized NBD-PtdCho was determined as above. In **f** and **g** the percentage of BSA-non-extractable NBD-PtdCho was determined in triplicate at 4 min (**f**) or 6 min (**g**) after the addition of NBD-PtdCho and is plotted as mean and s.d. All experiments were performed at least three times.

1–12, was identical in the patient and the parents, whereas its 3′ half, corresponding to exons 11–20, was shorter in the patient than in the parents (Fig. 4b). A sequence analysis indicated that the cDNA of the patient lacked the 226-bp sequence corresponding to exon 13. Direct sequencing of the chromosomal DNA indicated that the *TMEM16F* gene of the patient carried a G-to-T homozygous mutation at the splice-acceptor site in intron 12, whereas both parents were heterozygous for the mutation at this position (Fig. 4c). PCR analysis of the *TMEM16F* mRNA with primers at exons 12 and 16 showed a 608-bp band from the control and a 382-bp band from the cell line from the patient with Scott syndrome (Fig. 4d), indicating that a mutation in the splice acceptor site caused exon 13 to be skipped. This skipping caused a frame shift resulting in the premature termination of the protein in exon 14 (Fig. 4e) at the third transmembrane segment of human TMEM16F (Fig. 4f). The non-sense-mediated mRNA decay[21] may explain the decreased concentration of the exon-13-deleted form of *TMEM16F* mRNA in the patient's parents (Fig. 4d).

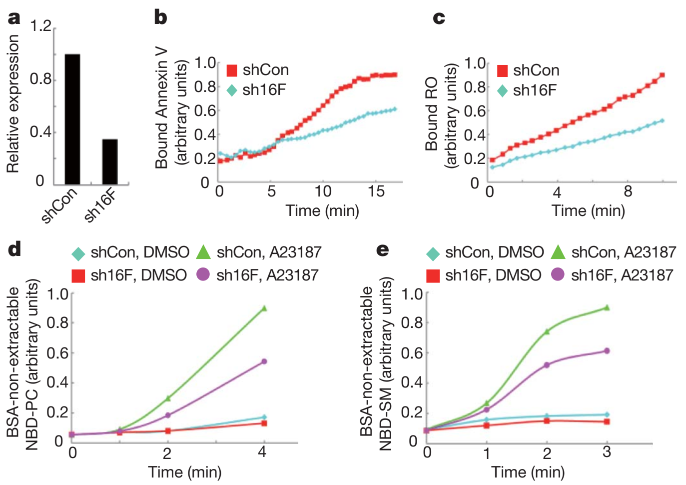


Figure 3 | Requirement of TMEM16F for phospholipid scrambling. **a**, Ba/F3 transformants expressing shRNA for *Tmem16f* (sh16F) or scrambled shRNA (shCon). *Tmem16f* mRNA level was normalized to *β-actin* mRNA and is shown as relative expression. **b**, **c**, Ba/F3 cells expressing sh16F or shCon were preincubated with Cy5-Annexin V (**b**) or biotin-Ro09-0198 (RO) and allophycocyanin (APC)-labelled streptavidin (**c**). A23187 was added and fluorescence was monitored. **d**, **e**, Ba/F3 cells expressing sh16F or shCon were preincubated with 0.5 µM NBD-PtdCho (**d**) or NBD-SM (**e**) in Hanks balanced salt solution containing Ca^{2+}. A23187 was added, incubated and diluted with fatty-acid-free BSA buffer, and fluorescence was determined. Experiments in **b**–**e** were performed at least three times.

Repeated FACS analysis has been used previously to establish cell lines that overexpress a particular cell-surface protein[22,23]. Here, this method yielded TMEM16F carrying a point mutation that rendered the process extremely sensitive to Ca^{2+}, such that in the cells expressing the mutated TMEM16F the phospholipid scramblase functioned even in resting cells, in which the cytosolic Ca^{2+} concentration was below 100 nM (ref. 24). The TMEM16 family, to which TMEM16F belongs, consists of ten members in humans and mice[5]. The founding member of the family, human TMEM16A, is a Ca^{2+}-dependent Cl^- channel[11–13]. Although the direct binding of Ca^{2+} to TMEM16 members has yet to be demonstrated, the amino-terminal region of TMEM16A seems to have a regulatory role[25]. Similarly, the increased sensitivity of the D409G mutant to Ca^{2+} suggests that either Ca^{2+} or a Ca^{2+}-sensing molecule binds to this N-terminal region of TMEM16F. The overexpression of TMEM16A in Ba/F3 cells had no effect on the ionophore-induced exposure of PtdSer (data not shown), suggesting that different members of this family have distinct functions. The PtdSer exposure or scrambling of phospholipids occurs in other biological processes[1,4,26–29], such as apoptotic cell death, the fusion of muscle, bone or trophoblast cells, and the release of neurotransmitters and microvesicles. It will be

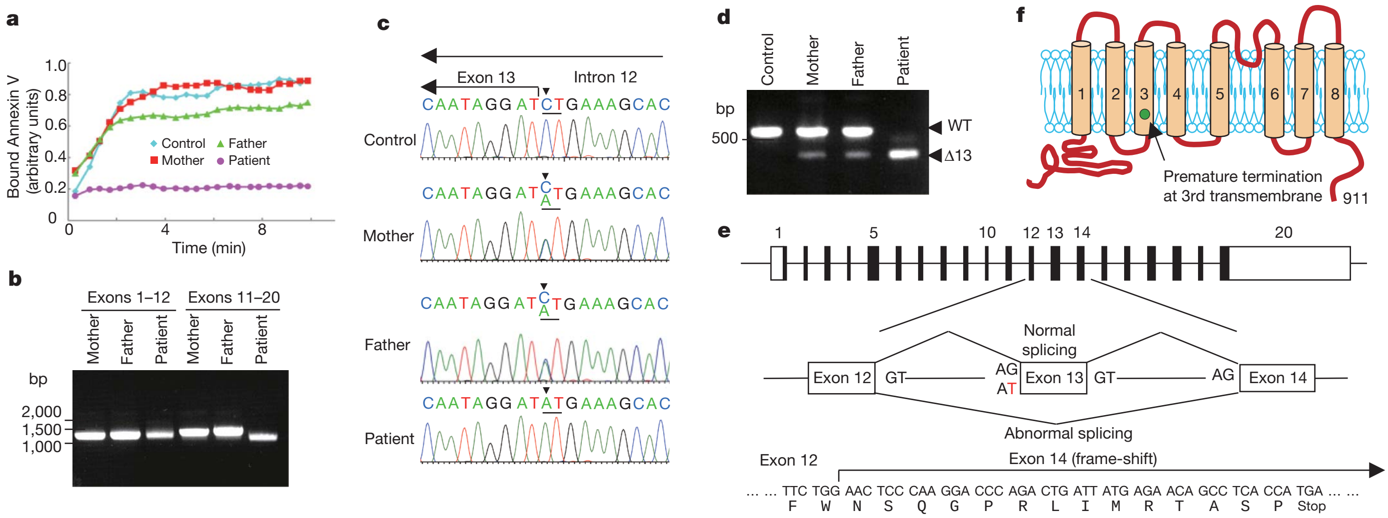


Figure 4 | A splice mutation of *TMEM16F* in a patient with Scott syndrome. **a**, Cells from control, from a patient with Scott syndrome and from the patient's parents, were preincubated with Annexin V. A23187 was added and fluorescence was monitored. **b**, RT–PCR for *TMEM16F* mRNA for exons 1–12 and 11–20 with RNA from the patient and parents. **c**, The junction between exon 13 and intron 12 of the *TMEM16F* gene sequenced from the 3′ end. The CT complementary to the splice acceptor site AG is underlined. Arrowheads indicate the mutation. **d**, RT–PCR for exons 12–16. Arrowheads indicate fragments for the wild-type (WT) and exon-13-deleted forms. **e**, Splicing in the patient's *TMEM16F* gene. The mutation causes the skipping of exon 13, resulting in a frame-shift mutation in exon 14. **f**, Schematic representation of the frame-shift mutation in the Scott patient that causes premature termination in the third transmembrane region of TMEM16F.

interesting to study whether TMEM16F and/or its related members in the TMEM16 family are involved in these processes.

METHODS SUMMARY

To expose PtdSer reversibly on the cell surface, Ba/F3 cells were treated at 37 °C with A23187 under Ca^{2+}-free conditions. The exposed PtdSer was detected by binding of Annexin V at 4 °C in Ca^{2+}-containing Annexin V-binding buffer. A subline (Ba/F3-PS19) of Ba/F3 cells that was extremely sensitive to Ca^{2+}-ionophore-elicited PtdSer exposure was selected by repeating the sorting 19 times with FACSAria (BD Bioscience). A cDNA library was established with mRNA from Ba/F3-PS19 cells in retrovirus vector, and the cDNA (*Tmem16f*) that caused Ba/F3 cells to expose PtdSer constitutively was identified by expression cloning. The Epstein–Barr virus (EBV)-transformed cell lines from a patient with Scott syndrome and from the patient's parents were described previously[20]. The *TMEM16F* mRNA in these cell lines was analysed by RT–PCR. The *TMEM16F* chromosomal gene was amplified by PCR from the genomic DNA of the cell lines, and was directly sequenced by cycle sequencing with an ABI 3100 genetic analyser (Applied Biosystems). Exposure of PtdSer and PtdEtn on the cell surface was analysed by the binding of Cy5-labelled Annexin V and biotin-labelled Ro09-0198 (ref. 17), respectively. The internalization of NBD-PtdCho and NBD-SM was analysed by the BSA-extraction method essentially as described[30]. For the knock-down experiment, shRNA-retrovirus vectors for *Tmem16f* and control scrambled sequence were obtained from OriGene, and the resultant retrovirus was used to infect Ba/F3 cells.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 April; accepted 13 October 2010.
Published online 24 November 2010.

1. Leventis, P. A. & Grinstein, S. The distribution and function of phosphatidylserine in cellular membranes. *Annu. Rev. Biophys.* **39,** 407–427 (2010).
2. Zwaal, R. F., Comfurius, P. & Bevers, E. M. Lipid–protein interactions in blood coagulation. *Biochim. Biophys. Acta* **1376,** 433–453 (1998).
3. Lentz, B. Exposure of platelet membrane phosphatidylserine regulates blood coagulation. *Prog. Lipid Res.* **42,** 423–438 (2003).
4. Balasubramanian, K. & Schroit, A. Aminophospholipid asymmetry: a matter of life and death. *Annu. Rev. Physiol.* **65,** 701–734 (2003).
5. Galietta, L. The TMEM16 protein family: a new class of chloride channels? *Biophys. J.* **97,** 3047–3053 (2009).
6. Weiss, H. & Lages, B. Family studies in Scott syndrome. *Blood* **90,** 475–476 (1997).
7. Zwaal, R., Comfurius, P. & Bevers, E. Scott syndrome, a bleeding disorder caused by defective scrambling of membrane phospholipids. *Biochim. Biophys. Acta* **1636,** 119–128 (2004).
8. Williamson, P. *et al.* Phospholipid scramblase activation pathways in lymphocytes. *Biochemistry* **40,** 8065–8072 (2001).
9. Wielders, S. J. *et al.* Absence of platelet-dependent fibrin formation in a patient with Scott syndrome. *Thromb. Haemost.* **102,** 76–82 (2009).
10. Daleke, D. Phospholipid flippases. *J. Biol. Chem.* **282,** 821–825 (2007).
11. Caputo, A. *et al.* TMEM16A, a membrane protein associated with calcium-dependent chloride channel activity. *Science* **322,** 590–594 (2008).
12. Schroeder, B., Cheng, T., Jan, Y. & Jan, L. Expression cloning of TMEM16A as a calcium-activated chloride channel subunit. *Cell* **134,** 1019–1029 (2008).
13. Yang, Y. *et al.* TMEM16A confers receptor-activated calcium-dependent chloride conductance. *Nature* **455,** 1210–1215 (2008).
14. Schreiber, R. *et al.* Expression and function of epithelial anoctamins. *J. Biol. Chem.* **285,** 7838–7845 (2010).
15. Shi, J. & Gilbert, G. Lactadherin inhibits enzyme complexes of blood coagulation by competing for phospholipid-binding sites. *Blood* **101,** 2628–2636 (2003).
16. Hanayama, R., Tanaka, M., Miwa, K. & Nagata, S. Expression of developmental endothelial locus-1 in a subset of macrophages for engulfment of apoptotic cells. *J. Immunol.* **172,** 3876–3882 (2004).
17. Emoto, K., Toyama-Sorimachi, N., Karasuyama, H., Inoue, K. & Umeda, M. Exposure of phosphatidylethanolamine on the surface of apoptotic cells. *Exp. Cell Res.* **232,** 430–434 (1997).
18. Macia, E. *et al.* Dynasore, a cell-permeable inhibitor of dynamin. *Dev. Cell* **10,** 839–850 (2006).
19. Toti, F., Satta, N., Fressinaud, E., Meyer, D. & Freyssinet, J. Scott syndrome, characterized by impaired transmembrane migration of procoagulant phosphatidylserine and hemorrhagic complications, is an inherited disorder. *Blood* **87,** 1409–1415 (1996).
20. Kojima, H. *et al.* Production and characterization of transformed B-lymphocytes expressing the membrane defect of Scott syndrome. *J. Clin. Invest.* **94,** 2237–2244 (1994).
21. Shyu, A., Wilkinson, M. & van Hoof, A. Messenger RNA regulation: to translate or to degrade. *EMBO J.* **27,** 471–481 (2008).
22. Kavathas, P. & Herzenberg, L. A. Amplification of a gene coding for human T-cell differentiation antigen. *Nature* **306,** 385–387 (1983).
23. Suda, T., Takahashi, T., Golstein, P. & Nagata, S. Molecular cloning and expression of the Fas ligand: a novel member of the tumor necrosis factor family. *Cell* **75,** 1169–1178 (1993).
24. Berridge, M., Bootman, M. & Roderick, H. Calcium signalling: dynamics, homeostasis and remodelling. *Nature Rev. Mol. Cell. Biol.* **4,** 517–529 (2003).
25. Ferrera, L. *et al.* Regulation of TMEM16A chloride channel properties by alternative splicing. *J. Biol. Chem.* **284,** 33360–33368 (2009).
26. Nagata, S., Hanayama, R. & Kawane, K. Autoimmunity and the clearance of dead cells. *Cell* **140,** 619–630 (2010).
27. Huppertz, B., Bartz, C. & Kokozidou, M. Trophoblast fusion: fusogenic proteins, syncytins and ADAMs, and other prerequisites for syncytial fusion. *Micron* **37,** 509–517 (2006).
28. Helming, L. & Gordon, S. Molecular mediators of macrophage fusion. *Trends Cell Biol.* **19,** 514–522 (2009).
29. van den Eijnde, S. *et al.* Transient expression of phosphatidylserine at cell–cell contact areas is required for myotube formation. *J. Cell Sci.* **114,** 3631–3642 (2001).
30. Williamson, P. *et al.* Transbilayer phospholipid movements in ABCA1-deficient cells. *PLoS ONE* **2,** e729 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Wiedmer for critical reading of our manuscript, and M. Fujii and M. Harayama for secretarial assistance. This work was supported in part by Grants-in-Aid for Specially Promoted Research (to S.N.) and for Young Scientists Start-up (to J.S.) from the Japan Society for the Promotion of Science. J.S. is supported by a research fellowship from the Japan Society for the Promotion of Science. P.J.S. was supported by grants from the National Institutes of Health, USA.

Author Contributions J.S. designed and performed the experiments, and wrote the manuscript. M.U. provided biotin-labelled Ro09-0198 peptide. P.J.S. provided EBV-transformed cell lines from a patient and the patient's parents, and commented on the manuscript. S.N. was responsible for the overall study design and for writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.N. (snagata@mfour.med.kyoto-u.ac.jp).

METHODS

Cell lines, recombinant proteins, antibodies, serum and reagents. Mouse interleukin (IL-3)-dependent Ba/F3 cells were maintained in RPMI medium containing 10% fetal calf serum (FCS; Gibco), 45 U ml^{-1} recombinant mouse IL-3 and 50 μM 2-mercaptoethanol. The EBV-transformed human cell lines[20] from a patient with Scott syndrome and the patient's parents were grown in RPMI1640 medium containing 10% FCS and 50 μM 2-mercaptoethanol. Human 293T cells and Plat-E packaging cells[31] were cultured in DMEM medium containing 10% FCS. Recombinant mouse IL-3 was produced by mouse C127I cells transformed with a bovine papillomavirus expression vector bearing mouse *Il-3* cDNA as described[32]. Biotin-labelled Ro09-0198 was prepared as described previously[33]. Flag-tagged mouse MFG-E8 was produced in human 293T cells as described[34], and the secreted MFG-E8 was purified with anti-Flag M2 beads (Sigma-Aldrich).

Ca^{2+}/Mg^{2+}-free RPMI1640 medium was purchased from Cell Science & Technology Institute. Ca^{2+}-free RPMI medium contained 0.5 mM $MgSO_4$. Ca^{2+}-free FCS was prepared by dialysing FCS for 2 days against PBS, with four changes of buffer. Dynasore was purchased from Calbiochem.

BAPTA-AM was from Dojindo. NBD-PtdCho and NBD-SM were purchased from Avanti Polar Lipids.

Treatment with Ca^{2+} ionophore, flow cytometry, and cell sorting. To expose PtdSer on the cell surface, 2×10^5 cells in a 96-well microtitre plate were washed with PBS, resuspended in 200 μl of HBSS (Gibco) and treated with A23187 (Sigma-Aldrich) at 37 °C for 15 min. The cells were stained on ice for 15 min with 2,500–5,000-fold diluted Cy5-labelled Annexin V (Biovision) in staining buffer (10 mM Hepes-NaOH pH 7.4 containing 140 mM NaCl and 2.5 mM $CaCl_2$) in the presence of 5 μg ml^{-1} PI. Flow cytometry was performed on a FACSAria (BD Bioscience) or FACSCalibur (BD Bioscience) and the data were analysed with FlowJo Software (True Star).

A subline of Ba/F3 cells that was sensitive to Ca^{2+}-ionophore-elicited PtdSer exposure was selected by repetitive sorting with a FACSAria. In brief, after 2×10^7 Ba/F3 cells in HBSS had been treated at 37 °C for 15 min with A23187, they were suspended in 1 ml of Annexin V staining buffer that had been prechilled to 4 °C. The cells were stained with Cy5-Annexin V on ice as described above, and sorted with a FACSAria whose injection chamber was kept at 4 °C. Cells providing the highest level of Cy5 fluorescence signal (the top 0.5–5.0%) were collected and resuspended at a density of more than 10^5 cells ml^{-1} in Ca^{2+}-free RPMI containing 5% dialysed FCS, 45 U ml^{-1} IL-3 and 50 μM 2-mercaptoethanol. After 24 h the cells were resuspended in normal Ca^{2+}-containing RPMI medium and expanded for the next sorting.

Construction of the cDNA library. Total RNA was prepared from Ba/F3 PS19 cells with an RNeasy Mini Kit (Qiagen), and poly(A)$^+$ RNA was purified with an mRNA Purification Kit (GE Healthcare) with two cycles of oligo(dT)-cellulose column chromatography. Double-stranded cDNA was synthesized with random hexamers as primers, using a cDNA synthesis kit (SuperScript Choice System for cDNA Synthesis; Invitrogen). A *Bst*XI adaptor was attached, and the fragments were size-fractionated by electrophoresis through a 1% agarose gel (Seakem GTG agarose; Lonza). DNA fragments longer than 2.5 kb were recovered from the gel with a DNA extraction kit (Wizard SV Gel and PCR Clean-up System; Promega) and ligated into a *Bst*XI-digested pMXs vector[35]. *Escherichia coli* DH10B cells (ElectroMax DH10B; Invitrogen) were transformed by electroporation with a Gene Pulser (Bio-Rad). About 9.3×10^5 clones were produced, and plasmid DNA was prepared with a QIAfilter Plasmid Maxi Kit (Qiagen).

Cell fusion. Ba/F3-PS0 and Ba/F3-PS19 cells were transduced with pMXs-puro EGFP and pMXs-neo DsRed, respectively, and cultured in the presence of 1 μg ml^{-1} puromycin or 1 mg ml^{-1} G418. Ba/F3-PS0 EGFP cells and Ba/F3-PS19 DsRed cells were fused in the presence of PEG1500 and cultured in the presence of 1 μg ml^{-1} puromycin and 1 mg ml^{-1} G418. The EGFP/DsRed double-positive cells were sorted with a FACSAria.

Screening of cDNA library. Plasmid DNA (108 μg) from the cDNA library was introduced by lipofection with FuGENE6 (Roche Diagnostics) into 7.2×10^7 PLAT-E packaging cells[31] grown in eighteen 10-cm dishes. Two days after the transfection, the viruses in the culture supernatant were centrifuged at 4 °C and 6,000*g* for 16 h, resuspended in RPMI1640 medium containing 10% FCS and 45 U ml^{-1} IL-3, and used to infect 7.2×10^6 Ba/F3 cells in the presence of 8 μg ml^{-1} Polybrene (Sigma-Aldrich). After a 24-h culture, the medium was replaced with fresh medium, and the cells were further cultured for 2 days. The sorting of cells that were sensitive to ionophore-induced PtdSer exposure was performed as described above.

Isolation of cDNA fragments from Annexin V-positive Ba/F3 cells. To isolate the cDNA integrated into the retroviral vector, the genomic DNA was extracted from Ba/F3 cell transformants with the Wizard Genomic DNA Purification System (Promega) and subjected to PCR with the Expand Long Template PCR System (Roche Diagnostics). The PCR primers (5′-CCCGGGGGTGGACCATCCTCT-3′ and 5′-CCCCTTTTTCTGGAGACTAAAT-3′) carried sequences from the pMXs vector, and the conditions for PCR were 10 s at 96 °C, 30 s at 58 °C and 4 min at 68 °C for 35 cycles. The PCR fragments were cloned into the pGEM-T Easy vector (Promega) and subjected to DNA sequencing analysis with an ABI PRISM 3100 Genetic Analyser (Applied Biosystems).

Expression vector for TMEM16F and its mutants. The Flag-tag sequence was integrated into the *Eco*RI and *Xho*I sites of the retroviral vector pMXs-puro, resulting in pMXs-puro c-Flag. The full-length coding sequence for mouse TMEM16F (GenBank accession number NM_175344) was prepared by RT–PCR with the mRNA from Ba/F3 cells. The primers used were as follows (in each primer the *Eco*RI recognition sequence is underlined): 5′-ATAT<u>GAATTC</u>GACATGCAGATGATGACTAGGAA-3′ and 5′-ATAT<u>GAATTC</u>GAGTTTTGGCCGCACGCTGT-3′.

The PCR fragments were inserted into the *Eco*RI site of pMXs-puro c-Flag, and the authenticity of the cDNAs was verified by DNA sequencing.

For the expression plasmid of TMEM16F–mRFP, the coding sequence for mRFP in pcDNA-mRFP (Invitrogen) was joined in-frame to the C terminus of mouse TMEM16F and introduced into pMXs vector.

Expression in mouse Ba/F3 and human 293T cells. The expression vector for Flag-tagged TMEM16F in pMXs-puro was introduced into Plat-E cells. The retrovirus produced was concentrated as described above and used to infect Ba/F3 cells to establish stable transformants. The transformants were selected by culturing the cells in medium containing puromycin (1.0 μg ml^{-1}). To express TMEM16F–mRFP, human 293T cells were transfected by lipofection with FuGENE6 with the pMXs vector carrying the TMEM16F–mRFP sequence. One day later, the transfected cells were observed by fluorescence microscopy (BioRevo BZ-9000; Keyence).

Western blotting. Cells were lysed in RIPA buffer (50 mM Hepes-NaOH pH 8.0 containing 1% Nonidet P40, 0.1% SDS, 0.5% sodium deoxycholate, 150 mM NaCl and 10% protease inhibitor cocktail (Complete Mini; Roche Diagnostics)). The lysate was mixed with 5 × SDS sample buffer (200 mM Tris-HCl pH 6.8, 10% SDS, 25% glycerol, 5% 2-mercaptoethanol, 0.05% bromophenol blue), boiled for 5 min and separated by electrophoresis on a 10% polyacrylamide gel (Bio Craft). After the proteins had been transferred to a poly(vinylidene difluoride) membrane (Millipore), the membranes were probed with horseradish peroxidase-conjugated mouse anti-Flag M2 (Sigma), and peroxidase activity was detected with a Western Lightning enhanced chemiluminescence system (PerkinElmer).

RT–PCR of *TMEM16F* cDNA and sequencing of its chromosomal gene in a patient with Scott syndrome. Total RNA was prepared from EBV-transformed cell lines from a patient with Scott syndrome and from the patient's parents, and from a healthy control. The RNA was reverse-transcribed with Superscript III (Invitrogen), in accordance with the manufacture's protocol, and the *TMEM16F* cDNA was analysed by PCR with the following sets of primers (in each primer the additional sequence is underlined): Ex1-FW (5′-<u>ATATGAATTC</u>GACATGAAAAGATGAGCAGGAA-3′), Ex11/12-RV (5′-<u>GCGTTCTTCTT</u>CCTGAGTAA-3′), Ex11/12-FW (5′-TTACTCAGGAAGAAGAACGC -3′), Ex20-RV (5′-<u>ATATGAATTC</u>TTCTGATTTTGGCCGTAAAT-3′), Ex12-FW (5′- TCTGTGCCAGTGCTGTCTTT-3′) and Ex16-RV (5′- CTGCAGATGGTAGTCCTGTT-3′).

For the sequence analysis of the human *TMEM16F* chromosomal gene, genomic DNA was prepared from human cell lines and a 965-bp DNA fragment carrying the 226-bp exon 13 and its 5′-flanking and 3′-flanking regions (about 370 bp each) was amplified by PCR with the following primers: 5′-CCAGAGTATGCTACTAGTTG-3′ and 5′-TCTCAGCAACCGAGGAACAT-3′. The PCR products were purified with a Wizard SV PCR and Gel Clean-up System. Cycle sequencing was performed with a BigDye Terminator v3.1 Cycle Sequencing kit with a primer of 5′-GGACCTTACCGAAGTTAGTA-3′, and analysed with an ABI PRIZM 3100 Genetic Analyser.

Analysis of exposure of PtdSer and PtdEtn. To analyse the exposure of PtdSer and PtdEtn, 10^5 cells at early exponential phase were washed with PBS, suspended in 1.0 ml of cold Annexin V staining buffer with 2,500–5,000-fold diluted Cy5-labelled Annexin V or 800-fold diluted biotin-Ro09-0198 (ref. 33) followed by 1.0 μg ml^{-1} APC-labelled streptavidin and 5 μg ml^{-1} PI. The samples were incubated on ice for 15 min, and flow cytometry was performed on a FACSAria or FACSCalibur as described above. For binding of MFG-E8, the cells were suspended in RPMI1640 containing 10% FCS and then incubated on ice for 20 min with Flag-tagged D89E mutant of MFG-E8 (0.4 μg ml^{-1})[34]. The cells were washed with the above medium and incubated on ice for 20 min with 1.0 μg ml^{-1} hamster monoclonal antibody against mouse MFG-E8 (clone 2422). This was followed by incubation with phycoerythrin-labelled mouse anti-hamster IgG (BD Bioscience) and analysis by flow cytometry with a FACSAria.

To study the requirement for intracellular Ca^{2+}, 10^5 cells were incubated with 10 μM BAPTA-AM in RPMI1640 medium containing 10% FCS at 37 °C for 5 min for the PtdSer exposure, or for 60 min for the PtdEtn exposure. The cells were

washed with Annexin V staining buffer, and stained with Cy5-Annexin V or biotin-Ro09-0198 as described above.

For the kinetic study of the Ca^{2+}-induced PtdSer and PtdEtn exposure, 10^6 cells were washed with PBS, suspended in 1.0 ml of cold Annexin V staining buffer with Cy5-labelled Annexin V or a mixture of biotin-Ro09-0198 and APC-labelled streptavidin, and 5 μg ml^{-1} PI. Cells were mixed on ice with A23187 at a final concentration of 0.25 or 0.5 μM, and applied to the injection chamber of a FACSAria that was set at 20 °C (for Ba/F3 cells) or 37 °C (for human cell lines) to induce the A23187 reaction. Data were recorded for the indicated periods, and the PI-positive cells were excluded from the analysis.

Internalization of NBD-PtdCho and NBD-SM. The internalization of NBD-lipid analogues was analysed by flow cytometry essentially as described in refs 8 and 30. In brief, 10^6 cells were washed with HBSS and resuspended in 0.5 ml of HBSS containing 2 mM $CaCl_2$ (HBSS-Ca). An equal volume of HBSS-Ca containing 1 μM NBD-PtdCho or NBD-SM was added to the cell suspension and incubated at room temperature. At each time point, 150 μl of cell suspension was collected, mixed with 150 μl of the prechilled (4 °C) HBSS-Ca containing 5 mg ml^{-1} fatty-acid-free BSA (Sigma-Aldrich), to extract the unincorporated fluorescent lipids, and 500 nM Sytoxblue (Molecular Probes). To measure the total fluorescence, samples were mixed with HBSS-Ca in the absence of BSA. After incubation for 10 min at 4 °C to extract the lipid, the cells were analysed with a FACSAria for forward scatter, side scatter, logarithmic green fluorescence (NBD), and Sytoxblue fluorescence. The Sytoxblue-positive dead cells were excluded from the analysis. The fluorescence of NBD-phospholipids that were resistant to the BSA extraction was regarded as representing phospholipids that had been incorporated into cells.

To examine the effect of the Ca^{2+} ionophore, 5×10^5 cells were washed with HBSS-Ca, resuspended in 0.5 ml of cold HBSS-Ca, and incubated on ice for 7 min. Cold HBSS (0.5 ml) containing 0.2 μM NBD-PtdCho or NBD-SM was added to the cell suspension and incubated further on ice for 3 min. The cells were then mixed with A23187 and incubated at room temperature to induce lipid incorporation. A 150-μl aliquot was used to determine the incorporated lipid quantity as described above.

Thin-layer chromatography. After incubation of cells with NBD-PtdCho or NBD-SM, the phospholipids were extracted from the cells by incubation at room temperature for 30 min with a mixture of chloroform, methanol and water (5:10:4, by volume). The phospholipids were separated by thin-layer chromatography on a silica gel 60 plate (Merck) with chloroform/acetone/methanol/acetic acid/water (5:2:1:1:0.5 by volume) as a solvent. The fluorescence on the plate was detected with a LAS4000 image analyser (Fuji Film).

Intracellular Ca^{2+} and Ca^{2+} influx. To determine the intracellular Ca^{2+} concentration, 10^6 cells were suspended in HBSS, incubated at 37 °C for 10 min with 0.4 μM Fluo-4-AM (Molecular Probes), washed with HBSS, and analysed with a FACSAria.

The Ca^{2+} influx was measured as described[36]. In brief, 10^6 cells were labelled for 30 min at 37 °C with 1 μM Fluo-4-AM in RPMI containing 10% FCS. After being washed with the Annexin V staining buffer, the cells were kept at 4 °C in Annexin V staining buffer. The Ca^{2+} ionophore A23187 was added to the mixture at a final concentration of 0.5 μM, and the change in mean fluorescence intensity was directly recorded with a FACSCalibur system. The data was analysed with FlowJo Software.

shRNA. shRNA expression plasmids for mouse *Tmem16f* in a pRS shRNA vector carrying the puromycin-resistance gene were purchased from OriGene. The target sequence of the shRNA for *Tmem16f* was 5′-CATCTACTCTGTGAAGTTC TTCATTTCCT-3′. The scrambled non-effective shRNA (5′-GCACTACCAGA GCTAACTCAGATAGTACT-3′) in pRS was from OriGene. Ba/F3 cells were infected with retrovirus containing the shRNA, and cultured in the presence of 1.0 μg ml^{-1} puromycin. Puromycin-resistant cells were subjected to cloning by limited dilution. The *Tmem16f* mRNA was quantified by real-time PCR, and the clones that showed the decreased expression were used for further study.

31. Morita, S., Kojima, T. & Kitamura, T. Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Ther.* **7,** 1063–1066 (2000).
32. Fukunaga, R., Ishizaka-Ikeda, E. & Nagata, S. Purification and characterization of the receptor for murine granulocyte colony-stimulating factor. *J. Biol. Chem.* **265,** 14008–14015 (1990).
33. Aoki, Y., Uenaka, T., Aoki, J., Umeda, M. & Inoue, K. A novel peptide probe for studying the transbilayer movement of phosphatidylethanolamine. *J. Biochem.* **116,** 291–297 (1994).
34. Hanayama, R. *et al.* Identification of a factor that links apoptotic cells to phagocytes. *Nature* **417,** 182–187 (2002).
35. Kitamura, T. *et al.* Retrovirus-mediated gene transfer and expression cloning: powerful tools in functional genomics. *Exp. Hematol.* **31,** 1007–1014 (2003).
36. Bernhagen, J. *et al.* MIF is a noncognate ligand of CXC chemokine receptors in inflammatory and atherogenic cell recruitment. *Nature Med.* **13,** 587–596 (2007).

doi:10.1038/nature09586

Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant *TET2*

Myunggon Ko[1]*†, Yun Huang[1]*†, Anna M. Jankowska[2], Utz J. Pape[1,3], Mamta Tahiliani[1], Hozefa S. Bandukwala[1], Jungeun An[1]†, Edward D. Lamperti[1], Kian Peng Koh[1], Rebecca Ganetzky[2], X. Shirley Liu[3], L. Aravind[4], Suneet Agarwal[5], Jaroslaw P. Maciejewski[2] & Anjana Rao[1]†

TET2 is a close relative of TET1, an enzyme that converts 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) in DNA[1,2]. The gene encoding TET2 resides at chromosome 4q24, in a region showing recurrent microdeletions and copy-neutral loss of heterozygosity (CN-LOH) in patients with diverse myeloid malignancies[3]. Somatic *TET2* mutations are frequently observed in myelodysplastic syndromes (MDS), myeloproliferative neoplasms (MPN), MDS/MPN overlap syndromes including chronic myelomonocytic leukaemia (CMML), acute myeloid leukaemias (AML) and secondary AML (sAML)[4–12]. We show here that *TET2* mutations associated with myeloid malignancies compromise catalytic activity. Bone marrow samples from patients with *TET2* mutations displayed uniformly low levels of 5hmC in genomic DNA compared to bone marrow samples from healthy controls. Moreover, small hairpin RNA (shRNA)-mediated depletion of *Tet2* in mouse haematopoietic precursors skewed their differentiation towards monocyte/macrophage lineages in culture. There was no significant difference in DNA methylation between bone marrow samples from patients with high 5hmC versus healthy controls, but samples from patients with low 5hmC showed hypomethylation relative to controls at the majority of differentially methylated CpG sites. Our results demonstrate that Tet2 is important for normal myelopoiesis, and suggest that disruption of TET2 enzymatic activity favours myeloid tumorigenesis. Measurement of 5hmC levels in myeloid malignancies may prove valuable as a diagnostic and prognostic tool, to tailor therapies and assess responses to anticancer drugs.

We transiently transfected HEK293T cells with Myc-tagged murine Tet2 and assessed 5mC and 5hmC levels by immunocytochemistry (Fig. 1 and Supplementary Figs 1–4). Myc–Tet2-expressing cells displayed a strong increase in 5hmC staining and a concomitant decrease in 5mC staining in the nucleus (Fig. 1b, c, quantified in Supplementary Fig. 4). In contrast, 5hmC was undetectable or barely detected in nuclei of cells expressing mutant Tet2 with H1302Y, D1304A substitutions in the signature HxD motif[1,12,17] involved in coordinating Fe^{2+}, and there was no obvious decrease in nuclear 5mC staining (Fig. 1b, c and Supplementary Fig. 4). These studies confirm that Tet2 is a catalytically active enzyme that converts 5mC to 5hmC in genomic DNA[13].

Mutations in TET2 residues H1881 and R1896, predicted to bind Fe^{2+} and 2-oxoglutarate (2OG), respectively, have been identified repeatedly in patients with myeloid malignancies[4,5,7,10]. HEK293T cells expressing Tet2 mutants H1802R and H1802Q (Fig. 1a and Supplementary Fig. 2) showed greatly diminished 5hmC staining and no loss of 5mC staining, consistent with participation of this residue in catalysis (Fig. 1b, c and Supplementary Fig. 4a, b). We analysed missense mutations identified in TET2 in our own (Supplementary Table 1) and other studies[3–6,11] (P1367S, W1291R, G1913D, E1318G and I1873T). HEK293T cells expressing Tet2 mutants P1287S, W1211R or C1834D (Supplementary Figs 2 and 3a) displayed low 5hmC staining and strong 5mC staining (Supplementary Figs 3b, c and 4c, d), indicating a role for these residues in the integrity of the catalytic or DNA binding domains. Cells expressing Tet2(R1817S/M) (Fig. 1a, Supplementary Figs 2 and 3a) were positive for 5hmC staining but changes in 5mC staining could not be assessed reliably (Fig. 1b, c, Supplementary Figs 3b, c and 4).

To quantify these findings, we developed dot blot assays to detect 5hmC in genomic DNA (Supplementary Fig. 5). In the first assay format, the blot was developed with a specific antiserum to 5hmC (Supplementary Fig. 5b, left), whose ability to recognize 5hmC depended strongly on the density of 5hmC in DNA (Supplementary Fig. 5c, top). We therefore developed a more sensitive and quantitative assay in which DNA was treated with bisulphite to convert 5hmC to cytosine 5-methylenesulphonate (CMS)[14] (Supplementary Fig. 5a), after which CMS was measured with a specific anti-CMS antiserum (Supplementary Fig. 5b, right). Unlike anti-5hmC which reacted efficiently only with DNA containing high densities of 5hmC, the anti-CMS antiserum recognized DNA with an average of only a single 5hmC per 201 base pairs (Supplementary Fig. 5c, bottom). This lack of density dependence allowed us to plot the signal obtained with twofold dilutions of a standard oligonucleotide containing a known amount of 5hmC against the amount of CMS obtained after bisulphite conversion. We assumed 100% conversion efficiency[15] and used the linear portion of the standard curve to compute the amount of CMS, and therefore 5hmC, in the DNA samples (for example, see Fig. 2a, right).

To assess 5hmC levels, we obtained uniform populations of Tet2-expressing HEK293T cells by transfection with Tet2-IRES-CD25 plasmid followed by magnetic isolation of CD25-expressing cells[1]. Wild-type and mutant Tet2 proteins were expressed at comparable levels (Fig. 1d and Supplementary Fig. 3d). Anti-5hmC/CMS dot blots of genomic DNA revealed, as expected, that 5hmC was barely detectable in DNA from cells transfected with empty vector; DNA from cells expressing wild-type Tet2 showed a substantial increase in 5hmC and a corresponding decrease in 5mC; and DNA from cells expressing the HxD mutant Tet2 protein had very low 5hmC (Fig. 1e, Supplementary Figs 3e and 6). DNA from cells expressing seven of the nine mutant Tet2 proteins tested—H1802Q/R, R1817S/M, W1211R, P1287S and C1834D—contained significantly less 5hmC than DNA from cells expressing wild-type Tet2 (Fig. 1e, Supplementary Figs 3e and 6), confirming our previous conclusion that these mutations impair enzymatic activity.

We measured 5hmC (CMS) levels in genomic DNA extracted from bone marrow or blood (with >20% immature myeloid cells) of 88

[1]Department of Pathology, Harvard Medical School, Immune Disease Institute and Program in Cellular and Molecular Medicine, Children's Hospital Boston, Boston, Massachusetts 02115, USA. [2]Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, and Department of Hematologic Oncology and Blood Disorders, Cleveland Clinic, Cleveland, Ohio 44195, USA. [3]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, Massachusetts 02115, USA. [4]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA. [5]Division of Pediatric Hematology/Oncology, Children's Hospital Boston and Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Boston, Massachusetts 02115, USA. †Present address: La Jolla Institute for Allergy and Immunology, La Jolla, California 92037, USA (M.K., Y.H., J.A., A.R.).
*These authors contributed equally to this work.

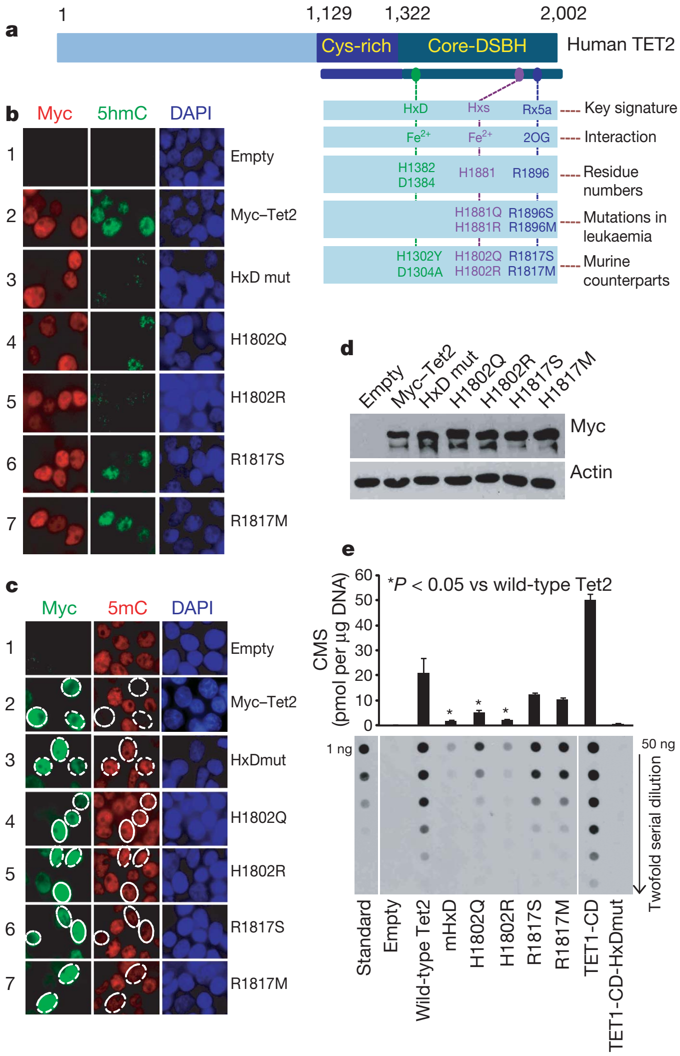


Figure 1 | The catalytic activity of Tet2 is compromised by mutations in predicted catalytic residues. **a**, Schematic representation of TET2. The catalytic core region contains the cysteine-rich (Cys-rich) and double-stranded beta-helix (DSBH) domains. Three signature motifs conserved among 2OG- and Fe^{2+}-dependent dioxygenases are shown[1,2]. Substitutions in the HxD signature that impair the catalytic activity of TET1 (ref. 1), leukaemia-associated mutations in the carboxy-terminal signature motifs, and corresponding substitutions introduced into murine Tet2 are indicated. **b**, Tet2 expression results in increased 5hmC by immunocytochemistry. HEK293T cells transfected with Myc-tagged wild-type and mutant Tet2 were co-stained with antibody specific for the Myc epitope (red) and antiserum against 5hmC (green). DAPI (blue) indicates nuclear staining. **c**, Tet2 expression results in loss of nuclear 5mC staining. HEK293T cells transfected with wild-type and mutant Myc-tagged Tet2 were co-stained with antibody specific for the Myc epitope (green) and antiserum against 5mC (red). **d**, Equivalent expression of wild-type and mutant Myc–Tet2. $CD25^{+}$ cells were isolated from HEK293T cells transfected with bicistronic Tet2-IRES-human CD25 plasmids, and Tet2 expression in whole cell lysates was detected by immunoblotting with anti-Myc. β-actin serves as a loading control. **e**, Genomic DNA purified from $CD25^{+}$ HEK293T cells overexpressing wild-type or mutant Tet2 was treated with bisulphite to convert 5hmC to CMS (Supplementary Fig. 5a). CMS was quantified by dot blot assay using anti-CMS and a synthetic bisulphite-treated oligonucleotide containing a known amount of CMS. As positive and negative controls, we included DNA from $CD25^{+}$ HEK293T cells transfected with TET1 catalytic domain (TET1-CD) or TET1-CD with mutations in the HxD motif (TET1-CD-HxDmut)[1].

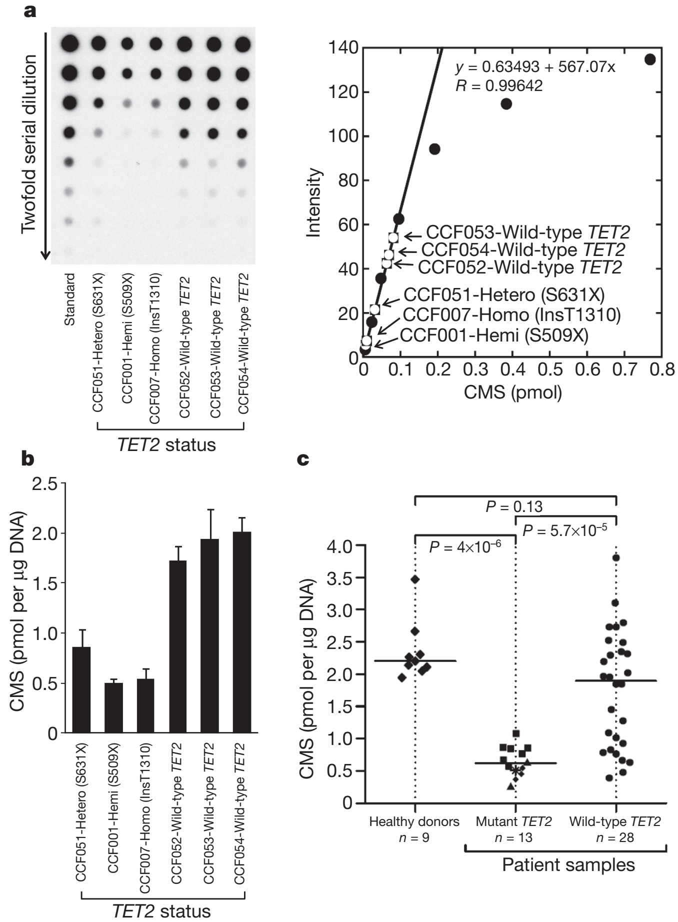


Figure 2 | TET2 mutational status correlates with 5hmC levels in patients with myeloid malignancies. **a**, Quantification of 5hmC by anti-CMS dot blot. Left, a representative dot blot of genomic DNA isolated from bone marrow aspirates of patients with MDS/MPN and *TET2* mutational status as indicated. A synthetic oligonucleotide with a known amount of CMS was used as standard. Right, the linear portion of the standard curve was used to estimate the amount of 5hmC in DNA from patient samples. **b**, Bar graph of data from panel **a**. The three patients with *TET2* mutations show lower 5hmC levels than the three patients with wild-type *TET2*. Error bars indicate s.d. ($n = 3$). **c**, Correlation of 5hmC levels with *TET2* mutational status. CMS levels in bone marrow samples from healthy donors and patients with myeloid malignancies (Supplementary Table 1) are shown as the median of triplicate measurements (Supplementary Fig. 7b). In the *TET2* mutant group, squares, triangles, diamonds and the star indicate homozygous, hemizygous, heterozygous and biallelic heterozygous mutations, respectively (for detailed definition, see Supplementary Methods). The horizontal bar indicates the median for each group. *P*-values for group comparisons were calculated by a two-sided Wilcoxon rank sum test. Patients bearing *TET2* mutations show uniformly low 5hmC expression levels.

patients with myeloid malignancies and 17 healthy controls (Supplementary Table 1). In blinded experiments, DNA was treated with bisulphite and CMS levels were measured. *TET2* mutations were strongly associated with low genomic 5hmC (Fig. 2 and Supplementary Fig. 7a). To confirm these conclusions in a statistically rigorous fashion, we tested samples for which a sufficient amount of DNA was available to make independent dilutions in triplicate, so that a median and standard deviation for 5hmC (CMS) levels in each patient could be derived (Supplementary Fig. 7b). Analysis of DNA from 9 healthy donors and 41 patients (28 with wild-type *TET2* and 13 with *TET2* mutations, Supplementary Table 1) revealed a strong, statistically significant correlation of *TET2* mutations with low 5hmC (Fig. 2c). In contrast, samples from patients with wild-type *TET2* showed a bimodal distribution, with 5hmC levels ranging from ~0.4 to ~3.8 pmol per µg DNA (Fig. 2c, Supplementary Fig. 7, also see Fig. 4).

We examined *Tet2* expression in haematopoietic cell subsets isolated from bone marrow and thymus of C57BL/6 mice (Supplementary Figs 8 and 9). *Tet2* mRNA was highly expressed in lineage-negative (Lin^{-}) $Sca\text{-}1^{+}c\text{-}Kit^{hi}$ (Sca-1 is also known as Ly6a) multipotent

progenitors (LSK), at levels similar to those in embryonic stem cells (ESC). Expression was maintained at high levels in myeloid progenitors (common myeloid progenitors, CMPs, and granulocyte-monocyte progenitors, GMPs), was low in mature granulocytes (Gr-1$^+$Mac-1$^+$, also known as Ly6g and Cd11b or Itgam, respectively) and high in monocytes (Gr-1$^-$Mac-1$^+$) (Supplementary Fig. 9a, middle panel).

To test the role of Tet2 in myelopoiesis, we transduced bone marrow stem/progenitor cells with *Tet2* shRNA (Supplementary Fig. 10a), effectively downregulating *Tet2* mRNA and protein relative to control cells transduced with empty vector or scrambled shRNA (Fig. 3a, b) (refer to Supplementary Fig. 10b for choice of *Tet2* shRNA). Tet2 depletion promoted expansion of Mac-1$^+$ F4/80$^+$ (also known as Emr1) and Mac-1$^+$ CD115$^+$ (also known as Csf1r or M-CSFR, macrophage colony stimulating factor receptor) monocyte/macrophage cells in the presence of G-CSF (granulocyte colony-stimulating factor) or GM-CSF (granulocyte-macrophage colony-stimulating factor), cytokines that support granulocyte and granulocyte/ monocyte development respectively, but not in the presence of M-CSF (macrophage colony-stimulating factor), which promotes growth of monocytic progenitors (Fig. 3c and Supplementary Fig. 10d). Simultaneous treatment with GM-CSF and M-CSF, or GM-CSF and G-CSF, also led to increased numbers of monocyte/macrophage cells (Fig. 3c). These results indicate that Tet2 has an important role in normal myelopoiesis. However, Tet2 does not markedly influence short-term proliferation of myeloid-lineage cells: when shRNA-transduced Lin$^-$ cells were cultured in the presence of GM-CSF and pulse-labelled with bromodeoxyuridine (BrdU), Tet2 depletion promoted monocyte/macrophage expansion but CD115$^+$ (M-CSFR$^+$) cells from the two cultures showed no difference in acute BrdU incorporation (Supplementary Fig. 11).

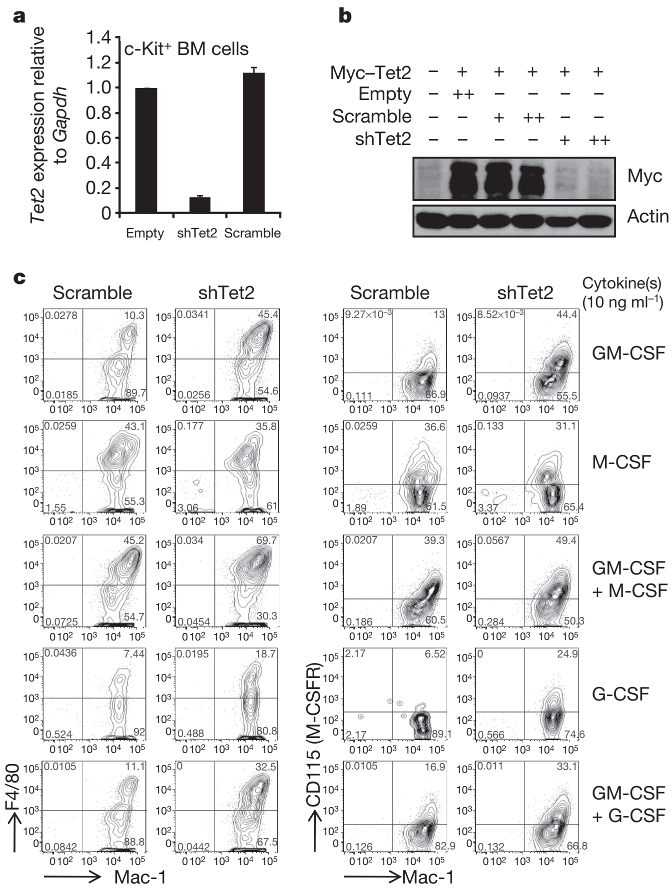


Figure 3 | Tet2 regulates myeloid differentiation. **a**, **b**, *Tet2* shRNA represses *Tet2* mRNA and protein expression. **a**, c-Kit$^+$ stem/progenitor cells from bone marrow of C57BL/6 mice were transduced with retroviruses (Supplementary Fig. 10). After selection with puromycin for 3 days, *Tet2* mRNA expression was assessed by quantitative RT–PCR (PCR with reverse transcription). Error bars show the range of duplicates. **b**, HEK293T cells were cotransfected with expression plasmids encoding Myc-tagged Tet2 and retroviral shRNAs. Tet2 protein expression was quantified 48 h later by anti-Myc immunoblotting of whole-cell extracts. **c**, Effect of Tet2 depletion on myeloid differentiation. Lin$^-$ cells purified from bone marrow of C57BL/6 mice were transduced with control (scramble) or shTet2 retroviruses, then grown in the presence of 50 ng ml^{-1} stem cell factor (SCF), puromycin (2 μg ml^{-1}) and cytokines (10 ng ml^{-1}) as indicated (also see Supplementary Fig. 10). After 4 days, flow cytometric analysis of Mac-1 versus F4/80 (left panel) or CD115 (right panel) was performed. All cells were GFP$^+$ on the day of analysis.

We asked whether 5hmC levels in tumour samples correlated with DNA methylation status. A histogram of normalized values from 88 patients and 17 healthy individuals showed the expected bimodal distribution (see Supplementary Methods): healthy controls and most patient samples with wild-type *TET2* had high 5hmC, whereas the majority of patient samples with mutant *TET2* had low 5hmC (Fig. 4b). The DNA methylation status of 62 samples was interrogated at 27,578 CpG sites. As expected[16], the resulting histograms were strikingly bimodal, with sites within and outside CpG islands showing low and predominantly high methylation, respectively (Fig. 4c). Comparison of 28 control samples with 24 high 5hmC tumour samples (22 wild-type *TET2*, 2 mutant *TET2*) showed no significant difference in DNA methylation; in contrast, comparison of the control samples with 29 low 5hmC tumour samples (7 wild-type *TET2* , 22 mutant *TET2*) yielded 2,512 differentially methylated sites, of which the majority (2,510 sites) were hypomethylated compared to controls (Fig. 4d and Supplementary Table 2). Thus TET2 loss-of-function is predominantly associated with decreased methylation at CpG sites.

To summarize, our studies demonstrate a strong correlation between myeloid malignancies and loss of TET2 catalytic activity. The leukaemia-associated missense mutations associated with diminished 5hmC levels provide clues to the structure of the TET2 catalytic domain. The W1211R, P1287S and C1834D mutations affect positions that are highly conserved within the catalytic domain of the TET subfamily of dioxygenases[2]: W1211 is located at the beginning of the strand just amino-terminal to the core of the double-stranded beta-helix (DSBH), and is predicted to constitute part of the 'mouth' of the active site pocket of the enzyme; P1287 is predicted to stabilize the conformation of the junction between the N-terminal helix and the first core strand of the DSBH; and G1913/C1834 is predicted to be the N-terminal capping residue of a helix that lines the 'mouth' of the DSBH and potentially interacts with substrate DNA[2]. The E1238G mutation had no detectable effect on 5hmC production in our overexpression assays; however, the patient with this mutation also showed CN-LOH spanning 4q24, a feature that likely contributes to the significant reduction in 5hmC levels observed in the bone marrow.

Low 5hmC levels were observed in a subset of patients with apparently wild-type *TET2*, whose clinical phenotypes resembled those of patients with mutant *TET2*. In several of these patients, *TET2* mRNA expression was not significantly different from controls; mutations in other TET proteins have not been described (Supplementary Text). Some patients in the wild-type *TET2*/low 5hmC category may harbour mutations in regulatory or partner proteins for TET2, or in *cis*-regulatory regions controlling *TET2* mRNA expression. Alternatively, the primary event in some of these patients may be CpG hypomethylation, resulting in decreased 5hmC secondary to depletion of the substrate, 5mC.

There is little consensus on whether *TET2* mutations correlate with clinical outcome. One study reported an association with decreased survival in AML[4], whereas others report little prognostic value in MPN diseases[7,10,12]. Assays for 5hmC may increase our options for the molecular classification of myeloid malignancies, making it possible to ask whether patients with high or low levels of genomic 5hmC show differences in disease progression or therapeutic response. Notably, histone deacetylase and DNA methyltransferase inhibitors show clinical efficacy

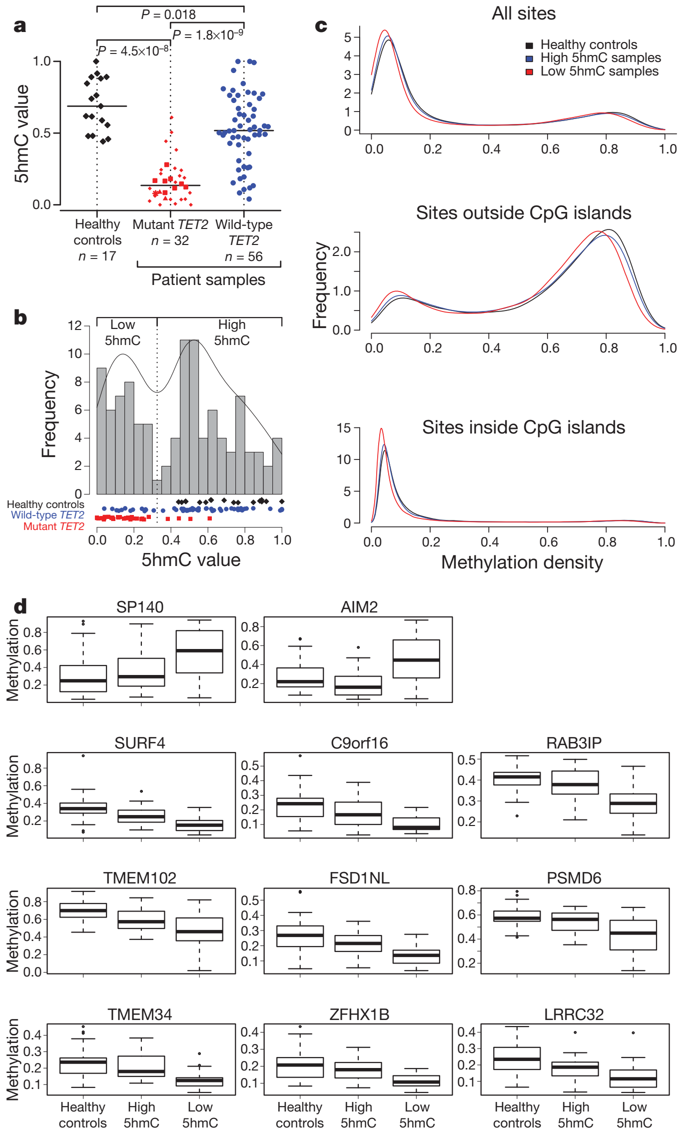


Figure 4 | Relation of 5hmC levels to DNA methylation status. **a**, Normalized 5hmC (CMS) levels in DNA from three different groups: healthy controls (black diamonds), patients with mutant *TET2* (red symbols) and patients with wild-type *TET2* (blue circles). Among *TET2* mutants, we distinguish homozygous (squares), hemizygous (triangles), heterozygous (small diamonds) and biallelic heterozygous (star) mutations (for definitions see Supplementary Methods). The horizontal bar indicates the median for each group. The number of samples in each group is indicated. **b**, Histogram of normalized 5hmC (CMS) levels in DNA from healthy donors (black diamonds), patients with mutant *TET2* (red rectangles) and patients with wild-type *TET2* (blue circles). The frequency was calculated based on a Gaussian kernel estimator. The local minimum between both modes was used as a threshold (vertical dotted line) between low and high 5hmC values. **c**, Density of methylation values for healthy controls (black), high 5hmC samples (blue) and low 5hmC samples (red) of all sites (top panel), sites outside CpG islands (middle panel) and sites inside CpG islands (lower panel). **d**, Box plot for group-specific methylation for the only two hypermethylated sites (*SP140*, *AIM2*; top panel) and the top nine hypomethylated sites (lower panels) between healthy controls and low 5hmC samples (total number of differentially methylated sites was 2,512).

in patients with CMML and AML[17]; and genomic 5hmC levels could potentially be a useful prognostic indicator or predictor of patient responses or refractoriness to 'epigenetic' therapy with demethylating agents.

DNA methylation is highly aberrant in cancer[18–20]. Because TET operates on 5mC, we were surprised to find that TET2 loss-of-function in myeloid tumours was associated with widespread hypomethylation rather than the expected hypermethylation at differentially-methylated CpG sites. Tumour samples with low 5hmC may have expanded cells with localized hypomethylation at these sites, or TET2 may control DNA methylation indirectly, for instance by regulating the expression or recruitment of one or more DNA methyltransferases, perhaps via 5hmC-binding proteins. Alternatively, if TET2 and 5hmC are required for cells to exit the stem cell state, loss of TET2 function in myeloid neoplasms may reactivate a stem-like state characterized by generalized hypomethylation and consequent genomic instability[21,22]. Indeed, hypomorphic *DNMT1* mutations associated with genome-wide DNA hypomethylation skew haematopoietic differentiation towards myelo-erythroid lineages[23], and promote the development of aggressive T-cell lymphomas due to activation and insertion of endogenous retroviruses[24,25]. Further studies of the role of TET2 in haematopoietic differentiation should uncover the relation between TET2 loss-of-function, DNA methylation changes and myeloid neoplasia.

METHODS SUMMARY

Patient samples. Genomic DNA was extracted from bone marrow/ peripheral blood samples from healthy donors and patients with MDS, MDS/MPN, primary and secondary AMLs. Clinical features and other detailed information pertaining to the patient samples are summarized in Supplementary Table 1.

Quantitative analysis of 5hmC and CMS levels using dot blot. For CMS detection, genomic DNA was treated with sodium bisulphite using the EpiTect Bisulfite kit (Qiagen). DNA samples were denatured and twofold serial dilutions were spotted on a nitrocellulose membrane in an assembled Bio-Dot apparatus (Bio-Rad). The blotted membrane was washed, air-dried, vacuum-baked, blocked and incubated with anti-5hmC or anti-CMS antibody (1:1,000) and horseradish peroxidase-conjugated anti-rabbit IgG secondary antibody. To ensure equal spotting of total DNA on the membrane, the same blot was stained with 0.02% methylene blue in 0.3 M sodium acetate (pH 5.2). To compare results obtained in different experiments, we used the normalization procedure described in Supplementary Methods (see Fig. 4a, b, which incorporate data from Fig. 2 and Supplementary Fig. 6).

Methylation analysis. The DNA methylation status of bisulphite-treated genomic DNA was probed at 27,578 CpG dinucleotides using the Illumina Infinium 27k array (Illumina)[26]. Methylation status was calculated from the ratio of methylation-specific and demethylation-specific fluorophores (β-value) using BeadStudio Methylation Module (Illumina). We removed sites on the Y and X chromosomes from the analysis because of inconsistent methylation status with respect to gender (a known problem based on communication with Illumina). Calculations are based on β values, which correspond to the methylation status of a site ranging from 0 to 1, returned by Illumina's BeadStudio software. We tested sites for differential methylation using an empirical Bayes approach employing a modified *t*-test (LIMMA). The false discovery rate (FDR) is controlled at a level of 5% by the Benjamini–Hochberg correction.

Received 29 March; accepted 19 October 2010.

Published online 7 November; corrected 9 December 2010 (see full-text HTML version for details).

1. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–935 (2009).
2. Iyer, L. M., Tahiliani, M., Rao, A. & Aravind, L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* **8,** 1698–1710 (2009).
3. Viguié, F. *et al.* Common 4q24 deletion in four cases of hematopoietic malignancy: early stem cell involvement? *Leukemia* **19,** 1411–1415 (2005).
4. Abdel-Wahab, O. *et al.* Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood* **114,** 144–147 (2009).
5. Delhommeau, F. *et al.* Mutation in *TET2* in myeloid cancers. *N. Engl. J. Med.* **360,** 2289–2301 (2009).
6. Jankowska, A. M. *et al.* Loss of heterozygosity 4q24 and *TET2* mutations associated with myelodysplastic/myeloproliferative neoplasms. *Blood* **113,** 6403–6410 (2009).
7. Langemeijer, S. M. *et al.* Acquired mutations in *TET2* are common in myelodysplastic syndromes. *Nature Genet.* **41,** 838–842 (2009).
8. Levine, R. L. & Carroll, M. A common genetic mechanism in malignant bone marrow diseases. *N. Engl. J. Med.* **360,** 2355–2357 (2009).
9. Mullighan, C. G. *TET2* mutations in myelodysplasia and myeloid malignancies. *Nature Genet.* **41,** 766–767 (2009).
10. Tefferi, A. *et al.* Frequent *TET2* mutations in systemic mastocytosis: clinical, *KITD816V* and *FIP1L1-PDGFRA* correlates. *Leukemia* **23,** 900–904 (2009).
11. Tefferi, A. *et al.* Detection of mutant *TET2* in myeloid malignancies other than myeloproliferative neoplasms: CMML, MDS, MDS/MPN and AML. *Leukemia* **23,** 1343–1345 (2009).
12. Tefferi, A. *et al.* *TET2* mutations and their clinical correlates in polycythemia vera, essential thrombocythemia and myelofibrosis. *Leukemia* **23,** 905–911 (2009).

13. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466,** 1129–1133 (2010).
14. Hayatsu, H. & Shiragami, M. Reaction of bisulfite with the 5-hydroxymethyl group in pyrimidines and in phage DNAs. *Biochemistry* **18,** 632–637 (1979).
15. Huang, Y. *et al.* The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS ONE* **5,** e8888 (2010).
16. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315–322 (2009).
17. Tefferi, A. Epigenetic alterations and anti-epigenetic therapy in myelofibrosis. *Leuk. Lymphoma* **49,** 2231–2232 (2008).
18. Smith, L. T., Otterson, G. A. & Plass, C. Unraveling the epigenetic code of cancer for therapy. *Trends Genet.* **23,** 449–456 (2007).
19. Esteller, M. Epigenetics in cancer. *N. Engl. J. Med.* **358,** 1148–1159 (2008).
20. Gal-Yam, E. N., Saito, Y., Egger, G. & Jones, P. A. Cancer epigenetics: modifications, screening, and therapy. *Annu. Rev. Med.* **59,** 267–280 (2008).
21. Ehrlich, M. DNA hypomethylation in cancer cells. *Epigenomics* **1,** 239–259 (2009).
22. Lengauer, C. Cancer. An unstable liaison. *Science* **300,** 442–443 (2003).
23. Bröske, A.-M. *et al.* DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nature Genet.* **41,** 1207–1215 (2009).
24. Gaudet, F. *et al.* Induction of tumors in mice by genomic hypomethylation. *Science* **300,** 489–492 (2003).
25. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genet.* **20,** 116–117 (1998).
26. Jiang, Y. *et al.* Aberrant DNA methylation is a dominant mechanism in MDS progression to AML. *Blood* **113,** 1315–1325 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by NIH grants R01 AI44432 and RC1 DA028422 (to A.R.), NIH grants K24 HL077522 and R01 HL098522, an Established Investigator award from the Aplastic Anemia & MDS Foundation, and an award from the Bob Duggan Memorial Research Fund (to J.P.M), NIH grant R01 HG4069 (to X.S.L.) and a pilot grant from Harvard Catalyst, The Harvard Clinical and Translational Science Center (NIH Grant #1 UL1 RR 025758-02, to S.A.). Y.H. was supported by postdoctoral fellowships from the GlaxoSmithKline-Immune Disease Institute (GSK-IDI) Alliance and the Leukemia and Lymphoma Society of America. H.S.B. is supported by a postdoctoral fellowship from the GSK-IDI Alliance.

Author Contributions M.K. analysed the biochemical effects of patient-associated TET2 mutations and performed the *in vitro* differentiation studies; Y.H. generated and characterized the anti-CMS antiserum, developed the quantitative dot-blot assay and quantified 5hmC in DNA samples from patients and healthy controls. A.M.J., R.G. and J.P.M. provided patient and control DNA for 5hmC quantification, performed DNA methylation arrays and analysed TET2 mutational status in patients. U.J.P. and X.S.L. carried out the statistical analysis of 5hmC levels and methylation data; M.T., H.S.B. and K.P.K. provided critical reagents; J.A. and E.D.L. contributed to molecular cloning and mouse maintenance respectively; and L.A. and S.A. provided essential intellectual input. A.R. set overall goals, coordinated collaborations and wrote the manuscript.

Author Information Data have been deposited at GEO under accession number GSE25706 (methylation status of each CpG site (beta value) can be found in Supplementary Table 3). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.R. (arao@idi.harvard.edu and arao@liai.org) or J.P.M. (maciejj@ccf.org).

doi:10.1038/nature09556

Design, function and structure of a monomeric ClC transporter

Janice L. Robertson[1], Ludmila Kolmakova-Partensky[1] & Christopher Miller[1]

Channels and transporters of the ClC family cause the transmembrane movement of inorganic anions in service of a variety of biological tasks, from the unusual—the generation of the kilowatt pulses with which electric fish stun their prey—to the quotidian—the acidification of endosomes, vacuoles and lysosomes[1]. The homodimeric architecture of ClC proteins, initially inferred from single-molecule studies of an elasmobranch Cl^- channel[2] and later confirmed by crystal structures of bacterial Cl^-/H^+ antiporters[3,4], is apparently universal. Moreover, the basic machinery that enables ion movement through these proteins—the aqueous pores for anion diffusion in the channels and the ion-coupling chambers that coordinate Cl^- and H^+ antiport in the transporters—are contained wholly within each subunit of the homodimer. The near-normal function of a bacterial ClC transporter straitjacketed by covalent cross-links across the dimer interface and the behaviour of a concatemeric human homologue argue that the transport cycle resides within each subunit and does not require rigid-body rearrangements between subunits[5,6]. However, this evidence is only inferential, and because examples are known in which quaternary rearrangements of extra-membrane ClC domains that contribute to dimerization modulate transport activity[7], we cannot declare as definitive a 'parallel-pathways' picture in which the homodimer consists of two single-subunit transporters operating independently. A strong prediction of such a view is that it should in principle be possible to obtain a monomeric ClC. Here we exploit the known structure of a ClC Cl^-/H^+ exchanger, ClC-ec1 from *Escherichia coli*, to design mutants that destabilize the dimer interface while preserving both the structure and the transport function of individual subunits. The results demonstrate that the ClC subunit alone is the basic functional unit for transport and that cross-subunit interaction is not required for Cl^-/H^+ exchange in ClC transporters.

To develop a strategy for generating a monomeric ClC protein, we examined the structure of ClC-ec1 (Fig. 1) for candidate residues mediating dimerization. This homologue is well suited to our purpose because its dimerization interface is almost completely membrane embedded, the large intracellular carboxy-terminal domain found in some ClC proteins being absent here. The interface is formed mainly by four helices running roughly perpendicular to the membrane to create a flat, nonpolar surface of ~1,200 Å^2 (Fig. 1). Most cross-subunit contacts are made by interdigitated leucine and isoleucine side chains; residues capable of forming hydrogen bonds or salt bridges are absent. The protein's phospholipid-facing residues are also nonpolar (Fig. 1), a circumstance that invites questions of how such chemically similar surfaces so faithfully choose their respective protein and lipid partners in the dimer. Such questions have motivated extensive studies of transmembrane peptide dimerization[8–10], which identified shape complementarity as an important determinant of helix packing specificity within membranes and micelles. Shape complementarity of the ClC-ec1 dimer interface is high, scoring at levels seen for protein-antibody contacts and several membrane protein oligomers (Supplementary Table 1). Accordingly, our design strategy seeks to destabilize the dimer by placing steric mismatches on the ClC subunit interface. A second element of the strategy aims at favouring the interface's exposure to the lipid bilayer. Lipid-facing surfaces of many membrane proteins are known to present amphiphilic tryptophan or tyrosine side chains to the chemically heterogeneous transition zone where the lipid acyl chains connect to the polar head groups, as seen for ClC-ec1 in Fig. 1; membrane-thermodynamic analysis of tryptophan analogues establishes that the aromatic, bifunctional character of this side chain favours its location at the phospholipid bilayer's transition zone[11].

With these considerations in mind, we adopted a 'warts-and-hooks' strategy for engineering a monomeric ClC by introducing tryptophan

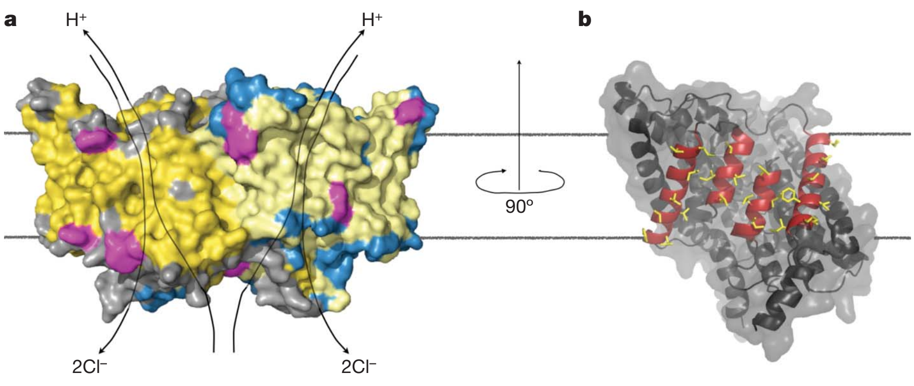


Figure 1 | Structure and dimeric interface of ClC-ec1. a, ClC-ec1 dimer (Protein Data Bank ID, 1OTS) is shown with subunits in grey and blue, with hydrophobic residues highlighted in yellow, and with tryptophan and tyrosine in magenta. The level of the membrane (extracellular side up) is indicated by black lines. Previously proposed transport pathways are shown for Cl^- and H^+. **b**, Single subunit rotated 90° to view the dimerization interface head-on. The four interface helices (residues 192–204, 215–232, 405–416 and 422–440) are shown in red and the side chains involved in cross-subunit contacts are shown in yellow.

[1]Department of Biochemistry, Howard Hughes Medical Institute, Brandeis University, Waltham, Massachusetts 02454, USA.

mutations on the subunit interface near the level of the lipid head groups. This type of substitution simultaneously offers two kinds of perturbation: steric disruption of the contact surface's shape complementarity and enhanced affinity of this surface for the lipid bilayer. We constructed eight single tryptophan substitutions for leucine or isoleucine near the extracellular and intracellular ends of the four dimerization helices (Fig. 2a). All but one of these mutants express near wild-type levels, and the oligomerization state of each was analysed in decylmaltoside micelles on a size exclusion column calibrated with a panel of membrane transport proteins[12] (Fig. 2b). The wild-type homodimer (100 kDa) elutes, as expected, at 12.8 ml, and the 50-kDa monomer is predicted to elute about 1 ml later. One mutant, Ile 422 Trp, shifts precisely to the presumed monomer position, with a minor dimer peak also apparent. Three other mutants, Ile 201 Trp, Leu 406 Trp and Leu 434 Trp, show broader, asymmetric peaks centred between dimer and monomer positions. The remaining three mutants all run as dimers (data not shown). In hopes of further stabilizing a monomer, we tested the double mutant Ile 201 Trp/Ile 422 Trp, which if dimeric would place four 'warts' within the subunit contact region, and if monomeric would offer two 'hooks' to the bilayer, one on each side of the membrane. This mutant, henceforth denoted WW, cleanly shifts to the monomer position with no observable dimer peak. The oligomeric nature of this double mutant in detergent micelles was further assessed by treatment with glutaraldehyde, a promiscuous crosslinker known quantitatively to produce covalent dimers of ClC-ec1[13], as illustrated for the wild type by SDS–polyacrylamide gel electrophoresis (Fig. 2c). In contrast, glutaraldehyde treatment fails to shift WW to the covalent-dimer position, thereby identifying it as a monomer in detergent.

To identify the oligomeric state of WW in lipid bilayers, we repeated glutaraldehyde crosslinking experiments on this protein reconstituted into liposomes. Phosphatidylcholine–phosphatidylglycerol mixtures were used here to avoid lipid-associated amino groups that would confound the glutaraldehyde reaction. We also aimed in these experiments to approximate Poisson-dilution conditions[14], wherein a low protein/lipid ratio is used so that most liposomes are protein free and any liposome containing protein carries only a single transporting unit. Under such conditions, each liposome becomes a single-molecule reaction vessel in which intramolecular crosslinking is favoured. As shown in Fig. 3a, crosslinking in liposomes recapitulates the detergent results, thereby showing that WW is also monomeric in these bilayer membranes.

The experiments above establish the WW mutant as monomeric but do not address its conformational or functional character. We therefore performed two mechanistically diagnostic ion-transport measurements in the same liposome environment as was used for the crosslinking experiments. The unitary passive Cl^- transport rate was determined in a 'Cl^- dump' experiment[14], in which liposomes with high Cl^- concentration are suspended in low-Cl^- solution in the presence of H^+ and K^+ ionophores, to prevent a pH gradient build-up and to maintain zero voltage. Under these conditions, the unitary Cl^- efflux rate of wild-type protein, measured electrochemically by the appearance of Cl^- in the external solution (Fig. 3b), is $\sim 300\ s^{-1}$; Cl^- turnover by WW is roughly half of this value ($160\ s^{-1}$), a respectable activity. Furthermore, anion specificity of transport is maintained in WW, as Cl^- efflux is fully dependent on addition of K^+ ionophore. ClC-ec1 is a coupled Cl^-/H^+ exchanger, in which a pre-established Cl^- gradient can be used to pump H^+ thermodynamically 'uphill'[15]. The WW monomer retains this defining feature of the transport

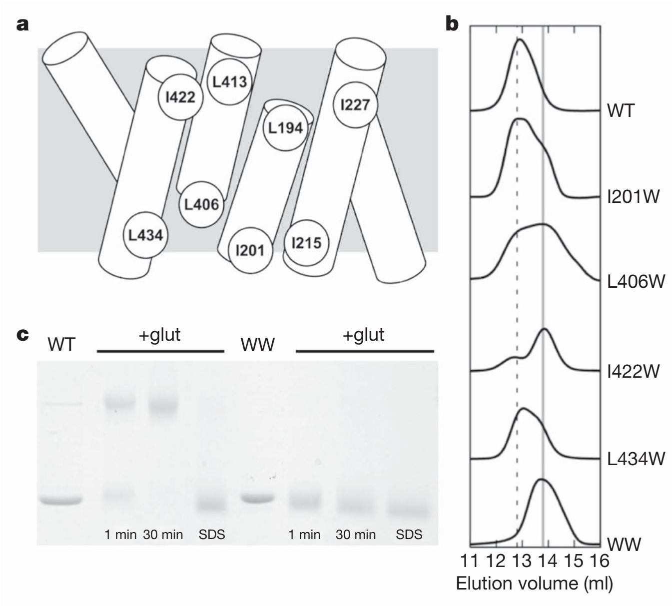


Figure 2 | Behaviour of tryptophan mutants in detergent. **a**, Schematic of the dimerization interface showing the positions of the tryptophans tested. Leu 194 Trp did not express protein. **b**, Chromatographic profiles of the various mutants on a Superdex 200 column. Vertical lines mark elution volumes for dimer (dashed) and monomer (solid). WT, wild type. **c**, 10% SDS–polyacrylamide gel electrophoresis of wild-type and WW samples, Coomassie stained. Bars indicate samples at 0.25 mg ml^{-1} treated with 0.125% glutaraldehyde, 150 mM NaCl and 50 mM Na phosphate, pH 7.0, for the indicated times in 5 mM decylmaltoside or, as a negative control, in 2% SDS. Crosslinking is nearly complete after 1 min, and no higher oligomers appear even after 30 min.

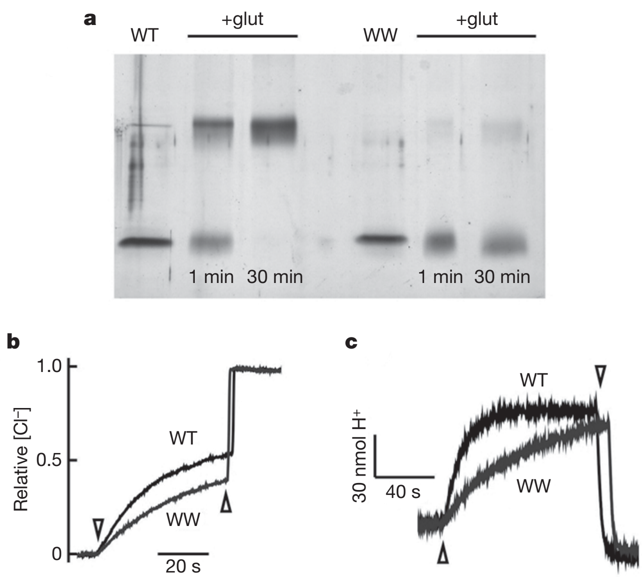


Figure 3 | Monomeric ClC mutant in phospholipid membranes. **a**, Glutaraldehyde crosslinking of wild-type ClC-ec1 and the WW mutant in liposomes. Glutaraldehyde treatment was as in Fig. 2, except that protein was incorporated into phosphatidylcholine–phosphatidylglycerol liposomes, and gel was silver stained. **b**, Passive Cl^- efflux from reconstituted liposomes for wild-type ClC-ec1 and the WW mutant. Traces show release of Cl^- from liposomes loaded with 300 mM Cl^- into the extraliposomal solution (containing 1 mM Cl^-), initiated by 0.5 μM valinomycin (downward arrowhead), normalized to the level of complete release on disrupting liposomes with 50 mM octylglucoside (upward arrowhead). Unitary turnover calculated on a per-subunit basis from the initial rate of Cl^- release[14] was $290 \pm 30\ s^{-1}$ for wild type, $160 \pm 9\ s^{-1}$ for WW (mean ± s.e.m., $N = 9$). [Cl^-], Cl^- concentration. **c**, Cl^--driven H^+ pumping against a pH gradient. Liposomes loaded with 300 mM Cl^-, pH 5.0, were suspended in 1 mM Cl^-, pH 5.2, and transport was initiated by valinomycin (upward arrowhead) and terminated by carbonyl cyanide-*p*-trifluoromethoxyphenylhydrazone (FCCP, downward arrowhead), while the pH of the suspension was recorded. Upward deflection represents uptake of H^+ into liposomes.

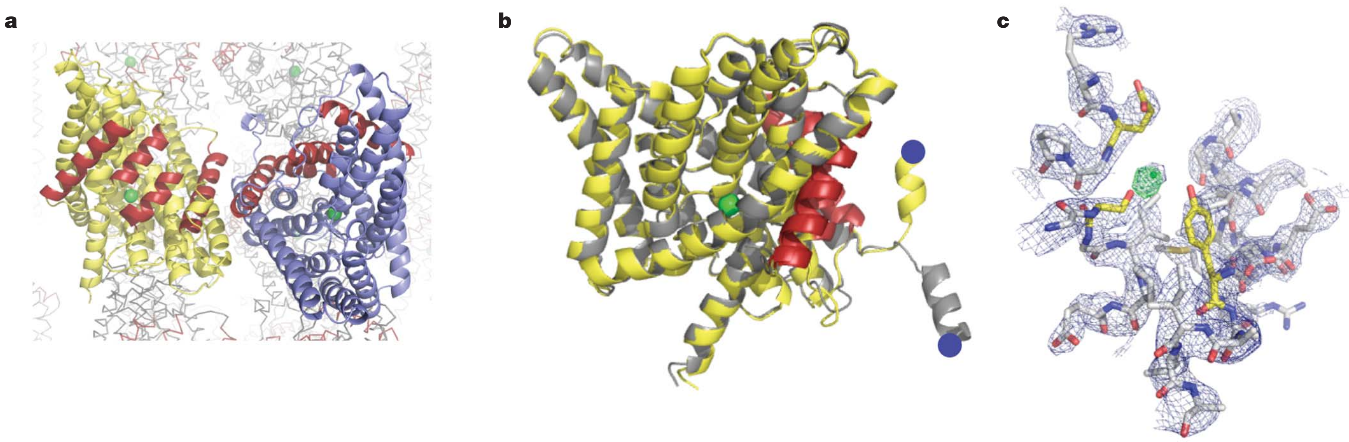


Figure 4 | Crystal structure of the WW monomer. **a**, View of two monomers in side-by-side contact, with interface helices highlighted in red, Cl^- ion highlighted in green and additional symmetry-related monomers shown in grey in the background. **b**, Backbone alignment (Cα root mean squared deviation, 0.6 Å) of the WW monomer (yellow, with interface helices in red) with a single subunit of wild-type ClC-ec1 (grey). Blue spheres indicate the N termini of the visible structures. **c**, Central anion-binding site. The $2F_o$–F_c map (blue, 1.5σ) is shown near the central Cl^--binding site, with coordinating residues Ser 107, Glu 148 and Tyr 445 highlighted (yellow); the positive difference density calculated from a Cl^--omit map (green) shows a strong peak (3.5σ) at the position of the central Cl^- ion in the wild type. Stereo versions of panels **a** and **c** can be found in Supplementary Fig. 2.

mechanism. As shown in Fig. 3c, Cl^--loaded liposomes are suspended in low-Cl^- medium, and transport is initiated by depolarizing the liposomes with K^+ ionophore. As Cl^- flows out, H^+ enters against a pH gradient, as detected by alkalinization of the extraliposomal medium, which is swiftly reversed by addition of a proton ionophore. We established Cl^-/H^+ exchange stoichiometry from the ratio of initial flux rates (Supplementary Fig. 1): 2.0 ± 0.1 for the wild-type control, as expected from the two-to-one stoichiometry determined in *E. coli* lipids[14–16], and a similar value, 2.3 ± 0.3, for WW. The preservation of H^+-coupled Cl^- antiport in the monomeric construct directly establishes that the ClC subunit contains all essential components of the transport mechanism. The possibility remains that side-chain movements at the dimer interface in wild-type homodimer may occur during the transport cycle, as indicated convincingly by recent ^{19}F NMR experiments[17], but our results demonstrate that such movements cannot represent functionally obligatory cross-subunit interactions.

We crystallized the WW mutant, collected X-ray diffraction data to a resolution of 3.1 Å and solved the structure by molecular replacement using the wild-type subunit as search model (crystallographic statistics are shown in Supplementary Table 2). The asymmetric unit consists of a single monomer whose previously buried dimer interface is now completely exposed to detergent-containing solvent. This exposed interface is shown in Fig. 4a (also see Supplementary Fig. 2) for a symmetry-related pair of monomers, whose contacts in the unit cell arise from crystal geometry and are not seen in crystals of wild-type ClC-ec1. We consider it remarkable that the monomer's 18 membrane-embedded helices align precisely with those of the wild-type subunit in the homodimer (Fig. 4b), despite the absence of native cross-subunit interactions. Only the cytoplasmic amino-terminal helix (residues 22–30), which in the wild type engages in a domain swap with its twin subunit, veers off in a different direction to accommodate crystal packing. Moreover, most side chains projecting from the exposed subunit interface are well ordered and unperturbed from their buried positions in the wild-type dimer, except for a single tyrosine, which adopts a different rotamer to make room for one of the substituted tryptophans (Supplementary Fig. 3). Unambiguous density for the mechanistically crucial central Cl^- ion appears in the monomer at the same position as in the wild type, coordinated by the central serine and tyrosine residues (Fig. 4 and Supplementary Fig. 2); however, Cl^- density is lower in WW than in wild-type data sets of similar crystallographic quality[18], perhaps because crystallization of the monomer requires the additional presence of NO_3^-, a transported anion known to compete with Cl^- (refs 13, 19).

A perplexing clash of form and function arises from this demonstration that the isolated ClC subunit is transport competent: why then are all known ClC proteins homodimers? With the steady expansion of the membrane protein structural database, it is becoming apparent that the parallel-pathways theme discussed here for ClCs appears in many families of channels and transporters. For instance, aquaporin channels are homotetramers with a diffusion pore in each subunit[20], FNT-family formate channels are five-pore pentamers[21,22] and UT-family urea channels[23], Amt-type ammonia channels[24,25] and outer-membrane porins[26] are three-pore trimers. Among membrane transporters, a striking example is found in five phylogenetically unrelated families whose transporting subunits share a common structural fold but variously assemble as monomers, dimers or trimers[27,28]. A survey of the current literature identifies no fewer than fourteen separate families (~40% of structurally known membrane transport protein families) built on this parallel-pathway principle, with subunits held together through extended, nonpolar intramembrane contacts. We are loath to offer any suggestion for the 'meaning'—evolutionary or physiological—of this emerging structural theme; in only one case, a trimeric Na^+-coupled aspartate transporter of the EAAT superfamily[29], has parallel-pathway architecture been plausibly proposed as essential for substrate transport by the individual subunits making up the complex.

Although our warts-and-hooks design succeeded in severing the ClC dimer, we do not claim to understand the thermodynamic reasons for its success. The energetic components governing how a greasy protein surface chooses its greasy protein partner over a greasy lipid bilayer are still unparsed. Previous attempts to attack this fundamental problem of membrane protein chemistry have focused on model systems of single transmembrane helical peptides[8–10,30], and most have been quantifiable only in detergent micelles. The ClC interface introduced here may provide future opportunities to examine the molecular forces operating in transmembrane helix packing, folding and recognition in the context of a complex integral membrane protein.

METHODS SUMMARY

Expression in *E. coli*, purification and liposome reconstitution of ClC-ec1 (Swiss-Prot ID, P37019) were performed as described[14], as were Cl^- and H^+ flux assays, except that we used lipid mixtures of egg phosphatidylcholine and 1-palmitoyl, 2-oleoyl phosphatidylglycerol in a 3/1 weight ratio. Ion flux rates in these lipids are 5–10-fold lower than observed in the *E. coli* phospholipids that we customarily use. Liposomes were formed at 20 mg ml^{-1} lipid, 1 μg protein per milligram lipid by dialysis, or by centrifugation of 0.1-ml samples through 3-ml Sephadex G-50 columns. Cl^-/H^+ exchange stoichiometry was determined as the ratio of initial

transport rates[16], with Cl^- efflux and H^+ uptake recorded by means of Cl^- and H^+ electrodes using liposomes loaded with 300 mM KCl and 40 mM citrate-NaOH, pH 5.0, suspended in solutions of 1 mM KCl, 300 mM K isethionate and 2 mM citrate-NaOH, pH 5.2. For crystallography, 1 μl ClC protein (9–15 mg ml^{-1}) in 100 mM NaCl, ~40 mM decylmaltoside and 10 mM Tris-HCl, pH 7.5, was mixed with an equal volume of 100 mM $LiNO_3$, 41–45% (w/v) PEG400, 100 mM glycine-NaOH, pH 9.5, and ~10 mM 4-cyclohexyl-1-butyl-β-D-maltoside was added to the 2-μl drop. Crystals grown by vapour diffusion in sitting drop trays for 2–4 weeks at 20 °C were frozen in liquid nitrogen, and data were collected remotely at beamline 8.2.1 of the Advanced Light Source Eastern Annex, Waltham, Massachusetts. Data were processed in HKL2000. Molecular replacement was done in PHASER using residues 30–450 of a single subunit of ClC-ec1 as search model, and refinement was carried out in REFMAC5.

Received 22 June; accepted 5 October 2010.
Published online 3 November 2010.

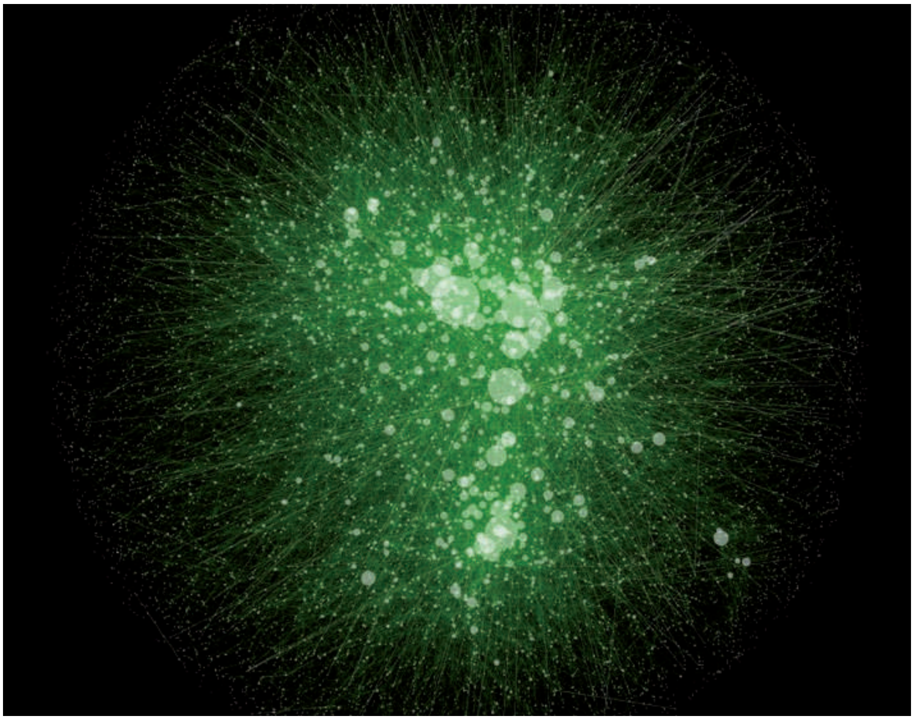
1. Jentsch, T. J. *et al.* Physiological functions of CLC Cl^- channels gleaned from human genetic disease and mouse models. *Annu. Rev. Physiol.* **67,** 779–807 (2005).
2. Middleton, R. E., Pheasant, D. J. & Miller, C. Reconstitution of detergent-solubilized Cl^- channels and analysis by concentrative uptake of $^{36}Cl^-$ and planar lipid bilayers. *Methods* **6,** 28–36 (1994).
3. Dutzler, R. *et al.* X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* **415,** 287–294 (2002).
4. Dutzler, R., Campbell, E. B. & MacKinnon, R. Gating the selectivity filter in ClC chloride channels. *Science* **300,** 108–112 (2003).
5. Nguitragool, W. & Miller, C. CLC Cl^-/H^+ transporters constrained by covalent cross-linking. *Proc. Natl Acad. Sci. USA* **104,** 20659–20665 (2007).
6. Zdebik, A. A. *et al.* Determinants of anion-proton coupling in mammalian endosomal CLC proteins. *J. Biol. Chem.* **283,** 4219–4227 (2008).
7. Bykova, E. A. *et al.* Large movement in the C terminus of CLC-0 chloride channel during slow gating. *Nature Struct. Mol. Biol.* **13,** 1115–1119 (2006).
8. Fleming, K. G., Ackerman, A. L. & Engelman, D. A. The effect of point mutations on the free energy of transmembrane α-helix dimerization. *J. Mol. Biol.* **272,** 266–275 (1997).
9. MacKenzie, K. R. & Fleming, K. G. Association energetics of membrane spanning α-helices. *Curr. Opin. Struct. Biol.* **18,** 412–419 (2008).
10. Chen, L. *et al.* Energetics of ErbB1 transmembrane domain dimerization in lipid bilayers. *Biophys. J.* **96,** 4622–4630 (2009).
11. Yau, W. M. *et al.* The preference of tryptophan for membrane interfaces. *Biochemistry* **37,** 14713–14718 (1998).
12. Fang, Y., Kolmakova-Partensky, L. & Miller, C. A bacterial arginine-agmatine exchange transporter involved in extreme acid resistance. *J. Biol. Chem.* **282,** 176–182 (2007).
13. Maduke, M., Pheasant, D. J. & Miller, C. High-level expression, functional reconstitution, and quaternary structure of a prokaryotic ClC-type chloride channel. *J. Gen. Physiol.* **114,** 713–722 (1999).
14. Walden, M. *et al.* Uncoupling and turnover in a Cl^-/H^+ exchange transporter. *J. Gen. Physiol.* **129,** 317–329 (2007).
15. Accardi, A. & Miller, C. Secondary active transport mediated by a prokaryotic homologue of ClC Cl- channels. *Nature* **427,** 803–807 (2004).
16. Miller, C. & Nguitragool, W. A provisional transport mechanism for a chloride channel-type Cl^-/H^+ exchanger. *Phil. Trans. R. Soc. B* **364,** 175–180 (2008).
17. Elvington, S. M., Liu, C. W. & Maduke, M. C. Substrate-driven conformational changes in ClC-ec1 observed by fluorine NMR. *EMBO J.* **28,** 3090–3102 (2009).
18. Accardi, A. *et al.* Synergism between halide binding and proton transport in a CLC-type exchanger. *J. Mol. Biol.* **362,** 691–699 (2006).
19. Nguitragool, W. & Miller, C. Uncoupling of a CLC Cl^-/H^+ exchange transporter by polyatomic anions. *J. Mol. Biol.* **362,** 682–690 (2006).
20. Fu, D. *et al.* Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **290,** 481–486 (2000).
21. Wang, Y. *et al.* Structure of the formate transporter FocA reveals a pentameric aquaporin-like channel. *Nature* **462,** 467–472 (2009).
22. Waight, A. B., Love, J. & Wang, D. N. Structure and mechanism of a pentameric formate channel. *Nature Struct. Mol. Biol.* **17,** 31–37 (2010).
23. Levin, E. J., Quick, M. & Zhou, M. Crystal structure of a bacterial homologue of the kidney urea transporter. *Nature* **462,** 757–761 (2009).
24. Khademi, S. *et al.* Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 Å. *Science* **305,** 1587–1594 (2004).
25. Zheng, L. *et al.* The mechanism of ammonia transport based on the crystal structure of AmtB of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **101,** 17090–17095 (2004).
26. Cowan, S. W. *et al.* Crystal structures explain functional properties of two *E. coli* porins. *Nature* **358,** 727–733 (1992).
27. Shaffer, P. L. *et al.* Structure and mechanism of a Na^+-independent amino acid transporter. *Science* **325,** 1010–1014 (2009).
28. Theobald, D. L. & Miller, C. Membrane transport proteins: surprises in structural sameness. *Nature Struct. Mol. Biol.* **17,** 2–3 (2010).
29. Reyes, N., Ginter, C. & Boudker, O. Transport mechanism of a bacterial homologue of glutamate transporters. *Nature* **462,** 880–885 (2009).
30. Metcalf, D. G. *et al.* Multiple approaches converge on the structure of the integrin αIIb/β3 transmembrane heterodimer. *J. Mol. Biol.* **392,** 1087–1101 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to the scientists at beamline 8.2.1 of the Advanced Light Source, Lawrence Berkeley Laboratory, for much help and advice, to H. Jayaram for help with refinement, and to R. Sah and D. Theobald for comments on the manuscript.

Author Contributions Experiments were designed by J.L.R. and C.M. and carried out by J.L.R. and L.K.-P., and the manuscript was written by all authors.

Author Information The atomic coordinates and structure factors of the WW (Ile 201 Trp/Ile 422 Trp) monomer have been deposited in the Protein Data Bank under accession code 3NMO. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.M. (cmiller@brandeis.edu).



K. ONO/UC SAN DIEGO/CYTOSCAPE

The human interactome contains more than 100,000 protein interactions, only a fraction of which are known.

Interactome under construction

Developing techniques are helping researchers to build the protein interaction networks that underlie all cell functions.

BY LAURA BONETTA

An old adage says: "Show me your friends, and I'll know who you are." In the same way, finding interaction partners for a protein can reveal its function. To that end, researchers are now building entire networks of protein–protein interactions. Unlike biological pathways, which represent a sequence of molecular interactions leading to a final result — for example, a signalling cascade — networks are interlinked. Represented as starbursts of protein 'nodes' linked by interaction 'edges' to form intricate constellations, they provide insight into the mechanisms of cell functions. Furthermore, placing proteins encoded by disease genes into these networks will let researchers determine the best candidates for assessing disease risk and targeting with therapies.

"This is the next step after the Human Genome Project," says Trey Ideker, a systems biologist at the University of California, San Diego, and principal investigator at the National Resource for Network Biology, which provides open-source software for network visualization. "That effort identified 30,000 genes, but that is not the end goal. How the genes work in pathways and how these pathways function in disease states and development is the end goal. To accomplish this we will need to systematically map gene and protein interactions."

Unlike the genome, the interactome — the set of protein-to-protein interactions that occurs in a cell — is dynamic. Many interactions are transient, and others occur only in certain cellular contexts or at particular times in development. The interactome may be tougher to solve than the genome, but the information, researchers say, is crucial for a complete understanding of biology.

THE RIGHT PARTNERS

At any time, a human cell may contain about 130,000 binary interactions between proteins[1]. So far, a mere 33,943 unique human protein–protein interactions are listed on BioGRID (http://thebiogrid.org), a database that stores interaction data. Clearly, there is work to do.

There are two main approaches for detecting interacting proteins: techniques that measure direct physical interactions between protein pairs — binary approaches — and those that measure interactions among groups of proteins that may not form physical contacts — co-complex methods (see 'Tools for the search').

The most frequently used binary method is the yeast two-hybrid (Y2H) system[2]. It has variations involving different reagents, and has been adapted to high-throughput screening. The strategy interrogates two proteins, called bait and prey, coupled to two halves of a transcription factor and expressed in yeast. If the proteins make contact, they reconstitute a transcription factor that activates a reporter gene.

Another method for identifying binary interactions is luminescence-based mammalian interactome mapping (LUMIER), a high-throughput approach developed by Jeff Wrana at the Samuel Lunefeld Research Institute in Toronto, Canada. This strategy fuses Renilla luciferaze (RL) enzyme, which catalyses light-emitting reactions, to a bait protein, which is expressed in a mammalian cell along with candidate protein partners tagged with a polypeptide called Flag. Researchers use a Flag antibody to immunoprecipitate all proteins with the Flag tag, along with any that interact with them. Interactions between the RL-fused bait and the Flag-tagged prey are detected when light is emitted. Other binary methods include the mammalian protein–protein interaction trap and techniques based on proteome chips.

The most common co-complex method is co-immunoprecipitation (coIP) coupled with mass spectrometry (MS). In this approach, a protein bait is tagged with a molecular marker. Several types of tags are commercially available; each requires a distinct biochemical technique to recognize the tag and fish the bait protein out of the cell lysate, bringing with it any interacting proteins. These are then identified by MS.

In addition to these empirical methods, researchers have used computational techniques to predict interactions on the basis of factors such as amino-acid sequence and structural information. "People ask 'Why are you predicting interactions when you can just do the experiment?'" says Gary Bader, a bioinformatician at the University of Toronto. "But experimental techniques fail for some proteins."

FALSE READINGS

Every step of a procedure to detect protein–protein interactions — from the reagents used to the cell types and experimental conditions — influences the proteins that are identified. Two studies this year used similar methods to identify interacting proteins in transcription factors in embryonic stem cells[3,4]; there was incomplete overlap between the resulting data sets. "If you use the same protocol you will get reproducible lists of proteins. But different labs use different protocols, which affects the end result," says Raymond Poot, a cell biologist at Erasmus MC

hospital in Rotterdam, the Netherlands, and lead author of one of the studies.

In his protocol, Poot pulled interacting proteins from cells using nuclear extracts expressing different Flag-tagged transcription factors. He added a nuclease to his reactions to remove DNA and eliminate possible artefacts caused by proteins binding to it. “Transcription factors bind to DNA so you are likely to pull out DNA-binding factors that are not directly interacting,” he explains. Purifying many different transcription factors with the same protocol also enabled the researchers to determine which interactions were most likely to be specific. For example, proteins that consistently co-purified with all transcription factors would be treated as unlikely to indicate a genuine interaction.

Calling out false positives — reported interactions that don’t actually occur — and false negatives — interactions that do occur but are not picked up by the experimental protocol or are discarded — is one of the main challenges in the field. “Normally when you do a coIP followed by MS you will get hundreds of protein candidates interacting with any one bait,” says Wade Harper, a cell biologist at Harvard Medical School in Boston, Massachusetts. “When you weed out all the stochastic and non-specific interactions you end up with many fewer proteins. Some proteins in large complexes might have 30–50 partners, others only 4–5.”

One way in which researchers increase the accuracy of their results is to use more than one method (for example, Y2H plus LUMIER) to detect the interactions. But the definition of a ‘real’ interaction depends on the context. “Does a real interaction mean that two proteins interact if they are placed next to each other in a test tube, or that they must interact in a cell? Or does real mean that the interaction should have a biological function?” asks Ideker. Researchers can home in on functional interactions by combining data on interactions with other types of biological information, such as genetic interactions, protein localizations or gene expression. For instance, proteins whose genes are co-expressed are likely to interact with each other or to be part of the same complex or pathway.

Many tools are available on the web for integrating different types of information about a given protein or gene. One is GeneMANIA, developed by Bader’s group in collaboration with Quaid Morris, a computational biologist also at the University of Toronto. A user enters the gene names into GeneMANIA; the program provides a list of genes that are functionally similar or have shared properties, such as similar expression or localization, and then displays a proposed interaction network, showing relationships among the genes and the type of data used to gather that information. The user can click on any node to obtain information about the gene and on any link to obtain information about their relationship (such as citations for any published studies or other sources of data). “It’s like a Google for genetic and protein information,” says Bader.

Other web-based interfaces that predict gene functions include STRING (http://string-db.org) developed at the European Molecular Biology Laboratory in Heidelberg, Germany. It hunts for protein interactions on the basis of genomic context, high-throughput experiments, co-expression and data from the literature.

Tools for the search

HYBRIGENICS

Methods such as the yeast two-hybrid system allow scientists to work out which proteins interact.

The two main methods for finding protein–protein interactions are the yeast two-hybrid (Y2H) system and co-immunoprecipitation followed by mass spectrometry. Several companies sell reagents for both approaches. Invitrogen of Carlsbad, California, sells the ProQuest Two-Hybrid System with Gateway Technology. This is based on Y2H, with modifications to decrease false-positive results and allow rapid characterization, says the company. Other firms provide vectors used to produce proteins with affinity tags, which can easily be immunoprecipitated along with other interacting proteins. A polypeptide tag called Flag is popular among researchers, and Sigma Aldrich of St Louis, Missouri, provides several Flag-genes for purchase. Promega in Madison, Wisconsin, has the HaloTag technology, in which a protein of interest is expressed in fusion with a tag protein engineered from a bacterial enzyme. This tag can be used to purify the protein, and any interacting with it, by binding to a resin. The tag is cleaved off using a protease.

For researchers who don’t have the time or infrastructure to do the experiments, companies such as Hybrigenics in Paris and Dualsystems Biotech of Schlieren, Switzerland, offer Y2H-based screening. “We have complex libraries with ten times more independent clones than most other libraries, which we screen to saturation. And rather than screening full-length proteins, we screen for interactions with domains,” says Etienne Formstecher, director of scientific projects and sales at Hybrigenics. “Full-length proteins can have some domains buried and not available to interact, at least in yeast where you may not have signals to unlock a closed protein conformation.” A customer is given a list of proteins that interact with the protein of interest; it indicates which domains are making contact and provides a confidence score for each interaction.

Innoprot in Derio, Spain, provides an interaction service using tag-based purification designed for high-throughput analysis. And Invitrogen’s ProtoArray Protein–Protein Interaction Service uses microarrays containing more than 9,000 human proteins to identify proteins that interact with any protein of interest. L.B.

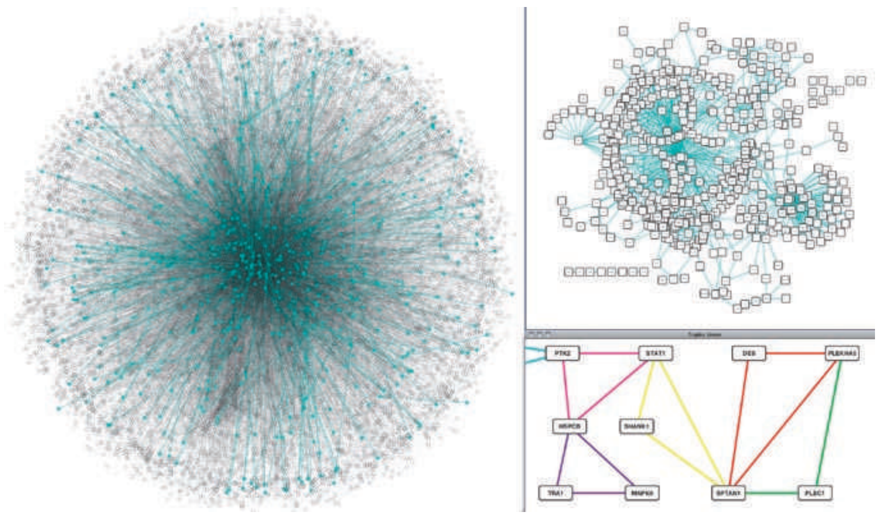
KEEPING SCORE

To select real protein–protein interactions, Harper and some members of his lab, Matt Sowa and Eric Bennett, developed a software platform called CompPASS to assign confidence scores to an interaction detected by MS[5]. CompPASS takes data sets of interacting proteins (including those identified in experiments) and measures frequency, abundance and reproducibility of interactions to calculate the score.

This year, Harper used CompPASS to identify interactions among proteins involved in autophagy, the process by which cellular proteins and organelles are engulfed into vesicles and delivered to the lysosome to be degraded. Starting with 32 proteins known to have a role in autophagy, they identified 2,553 interacting proteins using coIP–MS. CompPASS then narrowed the list down to 409 high-confidence interacting proteins with 751 interactions[6].

Ideker’s group used a different approach to map interactions among human mitogen-activated protein kinases (MAPKs), which respond to external stimuli and regulate cell function. Having used Y2H to identify more

K. ONO/UC SAN DIEGO/CYTOSCAPE



From a full network, researchers can zoom in on specific interactions that might be functionally relevant.

than 2,000 interactions among known MAPKs, Ideker used evidence including conservation of interactions among different species to winnow that down to a core network of 641 high-confidence interactions[7].

For some of the proteins there was no previous evidence of interactions with MAPKs. Ideker and his colleagues knocked down the expression of these proteins using RNA interference, then looked for the effect of the knockdowns on proteins known to be activated by MAPKs. This allowed them to confirm that about one-third of their interactions had a role in MAPK signalling.

These methods are helping to weed out false positives and provide associated confidence scores, but the problem of false negatives persists. "With these assays we try to get false positives down to zero. The hit you take is on false negatives. So now you can be highly confident of your data but you are probably probing only about 20% of the interactome," says Ideker. "We would like to get every interaction but we do not get even close with current technologies."

New methods may become available to identify interactions that escape detection by current techniques (see 'Real-time analysis'). In the meantime, one way to address the problem is to combine procedures for detecting interactions, each sampling a different portion of the interactome. The interaction data obtained in an experiment can also be combined with that available in public databases, thus providing a more complete picture, says Bader.

FROM DATA TO NETWORKS

Protein–protein interactions are only the raw material for networks. To build a network, researchers typically combine interaction data sets with other sources of data. Primary databases that contain protein–protein interactions include DIP (http://dip.doe-mbi.ucla.edu), BioGRID, IntAct (www.ebi.ac.uk/intact) and MINT (http://mint.bio.uniroma2.it). These databases have committed to making records available through a common language called PSICQUIC, to maximize access.

Other types of data that can be combined with protein–protein interactions include information on gene expression, cellular co-localization of proteins (based on microscopy), genetic information, metabolic and signalling pathways, and data from high-throughput assays.

"One challenge computationally is integrating heterogeneous data sets to build a network model," says Ilya Shmulevich, a professor at the Institute for Systems Biology in Seattle, Washington. The second challenge is to decide on a modelling approach. "It will depend on what kind of data you have available and how you will be using the model," says Shmulevich.

Several bioinformatic tools have been developed to model and represent networks. The most widely used ones are associated with Cytoscape (www.cytoscape.org), an open-source program for visualizing networks and for integrating them networks with other types of data. Several Cytoscape plug-ins allow users to download and explore databases.

Commercial packages with similar functions include MetaCore from GeneGO in St Joseph, Michigan; Pathway Analysis from Ingenuity Systems in Redwood City, California; and Pathway Studio from Ariadne Genomics in Rockville, Maryland. These can access public sources of data as well as the company's proprietary databases. "One of the unique features of Pathway Studio is the openness of our system and the ability to integrate many different kinds of data," says David Denny, director of marketing and product management at Ariadne.

GUILT BY ASSOCIATION

One reason for developing networks is to help assign functions to proteins through guilt by association. But "a huge slice of the proteome consists of proteins that no one knows what they do or interact with", says Benjamin Cravatt, a chemical physiologist at the Scripps Research Institute in San Diego, California.

For proteins not yet assigned to a portion of the human interaction network, Cravatt's group developed a technology for assigning protein functions by exploiting an interaction between enzymes and chemical reagents dubbed activity-based probes. These probes consist of a reactive group that binds the active sites of many members of an enzyme family, and a reporter tag that is used for the detection and identification of the probe-labelled enzymes[8].

Because these probes bind only to enzymes that are active, they can give insights into the enzymes' functions. For example, if a probe binds to a set of enzymes in a cancer cell but not in a normal cell, it means that these enzymes become more active in the cancer cell and so may have a role in cell growth. The activity probes can also serve as assays for the discovery of inhibitors for a particular enzyme, which may help researchers to understand the role of that enzyme. "You can develop an inhibitor for an enzyme before ever knowing what the actual substrate is," says Cravatt.

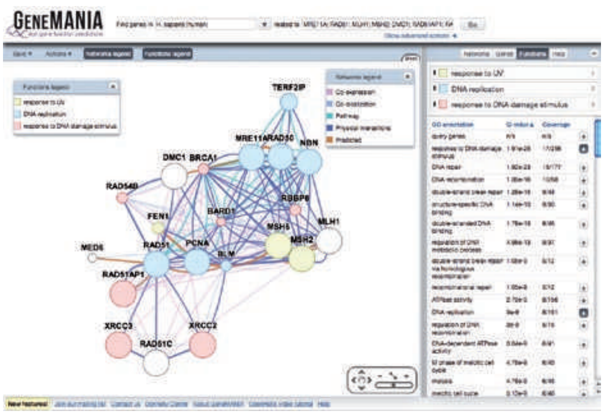
This year, he developed another strategy that not only determines differences in enzyme activities in different cells, but also pinpoints where in the protein these differences occur, providing a more quantitative measure of the differences[9]. The activities of many families of enzymes are regulated or fine-tuned by cysteine modifications. By looking specifically for changes in cysteine modifications across the proteome, he found 'hyper-reactive' cysteine residues in several proteins of unknown function, which suggests that they probably have roles in signalling pathways.

"One challenge is integrating heterogeneous data sets."
Ilya Shmulevich

One challenge in defining protein–protein interaction networks is that interactions vary depending on the type of cell and the cellular environment. For example, Wrana mapped the protein–protein interaction network for TGF-β, a growth factor that regulates cell functions, and found that two proteins that pass on the signals from the factor inside the cell — Smad2 and Smad4 — interact with one another only when the cells are stimulated with TGF-β. If the cells are not stimulated, these two proteins don't come into contact[10].

Bennett, Harper and Steven Gygi, a cell biologist also at Harvard Medical School, developed a proteomics platform centred around a technology called multiplex absolute quantification (AQUA) to look at dynamic changes in protein interaction networks. AQUA uses synthetic peptides that contain stable isotopes as internal standards for the native peptides that are produced when proteins from a cell

lysate are digested. Using tandem MS, researchers can compare the levels of native and synthetic peptides in a cell to obtain a measure of the amount of native proteins present. Synthetic peptides can also be prepared with modifications, such as extra phosphate groups, to measure the number of post-translationally modified proteins. "We are pursuing the dynamics of protein networks by quantifying changes in the amount of proteins present in specific protein complexes," says Harper. "Techniques such as AQUA provide an accurate and sensitive measure of how the stoichiometry of components within complexes that make up a network are altered in response to a stimulus."



Web tools such as GeneMANIA integrate data on a protein or gene.

The team used the approach to describe the rearrangements that occur in the protein network of cullin-RING ubiquitin ligases, enzymes that regulate protein turnover, under various cellular conditions[11].

DEVELOPMENTS IN DIAGNOSTICS

Changes in protein–protein interaction networks may provide information about the mechanisms of disease. Last year, Wrana applied the network approach to the diagnosis of breast cancer. He used microarrays to measure genome-wide protein expression in the tumours of people with breast cancer, and then overlaid the expression data on the network diagram of the human interactome.

Wrana had noted that 'hub' proteins, defined as those that interact with more than four others, can be grouped into two categories depending on whether they are expressed at the same time as the proteins with which they interact. When they looked at breast-cancer samples, Wrana and his colleagues found that certain hub proteins were in a different category in breast-cancer patients with a good prognosis than in those with a poor prognosis.

Thus, by overlaying the expression pattern of a cancer cell from an individual patient onto the human interactome network, Wrana could predict a patient's prognosis. "We found that the detection of global changes in network organization is more predictive of outcome than is gene expression alone," says Wrana. "We have now applied this method to other tumour models and obtained similar results."

KAYAK (kinase activity assay for protein profiling) is another approach to developing diagnostic tools for cancer on the basis of the functional consequences of the interaction between a protein, in this case a kinase, and its substrate. In this method, up to 90 peptide substrates for kinases are used to simultaneously measure the addition of phosphate groups to proteins in a cell lysate — in essence providing a 'phosphorylation signature' for that particular cell. "The readout is so sensitive and so quantitative that even small differences are teased out," says Gygi, who helped to develop the method[12].

According to Gygi, the biggest application of KAYAK might be in tumour classification. "Biopsies or excised tissues can be profiled for kinase activities with pinpoint accuracy. These patterns could contribute towards personalized drug treatments based on dysregulated kinase pathways," he says.

The combination of different types of data and technologies should continue to fill in the empty spaces of the current human interactome map. The picture may never be complete, but it will continue to provide insights into cellular mechanisms of health and disease. "I think that the network we have is dense enough for us to start doing studies to classify disease states," says Wrana. "As the networks become better and coverage improves, the accuracy of diagnosis will also improve." ■

Laura Bonetta *is a freelance science writer based in Garrett Park, Maryland.*

1. Venkatesan, K. *et al. Nature Meth.* **6,** 83–90 (2009).
2. Fields, S. & Song, O.-K. *Nature* **340,** 245–246 (1989).
3. van den Berg, D. L. C. *et al. Cell Stem Cell* **6,** 369–381 (2010).
4. Pardo, M. *et al. Cell Stem Cell* **6,** 382–395 (2010).
5. Sowa, M. E., Bennett, E. J., Gygi, S. P. & Harper, J. W. *Cell* **138,** 389–403 (2009).
6. Behrends, C., Sowa, M. E., Gygi, S. P. & Harper, J. W. *Nature* **466,** 68–76 (2010).
7. Bandyopadhyay, S. *et al. Nature Meth.* **7,** 801–805 (2010).
8. Nomura, D. K., Dix, M. M. & Cravatt, B. F. *Nature Rev. Cancer* **10,** 630–638 (2010).
9. Weerapana, E. *et al. Nature* **468,** 790–795 (2010).
10. Barrios-Rodiles, M. *et al. Science* **307,** 1621–1625 (2005).
11. Bennett, E. J. *et al. Cell* (in the press).
12. Yonghao, Y. *et al. Proc. Natl Acad. Sci. USA* **106,** 11606–11611 (2009).
13. Uemura, S. *et al. Nature* **464,** 1012–1017 (2010).

Real-time analysis

In November, Pacific Biosciences of Menlo Park, California, commercially released its third-generation DNA-sequencing platform, based on its single-molecule, real-time (SMRT) technology. A single DNA polymerase bound to a DNA template is attached to a tiny chamber illuminated by lasers, and nucleotides labelled with coloured fluorophores are introduced to it. As the polymerase incorporates them, each base is held for a few microseconds, while the fluorophore emits coloured light corresponding to the base identity. SMRT technology could also be used to analyse biomolecules other than DNA, and could become a common tool for detecting protein interactions, with some unique features. "This technology can detect relatively weak interactions," says Jonas Korlach, a scientific fellow at Pacific Biosciences, adding that it could pick out interactions that happen so quickly that they can't be identified by current methods.

As a step towards such applications, Joseph Puglisi, a structural biologist at Stanford University School of Medicine in California, and his group, with scientists at Pacific Biosciences, observed transfer RNAs binding to single ribosomes in real time[13]. In an unpublished follow-up, Puglisi's group has used SMRT technology to watch interactions between transfer RNAs, ribosomes and protein factors to determine how the translation machinery synthesizes proteins. "We have just seen the tip of the iceberg in terms of applications," says Korlach. L.B.

PACIFIC BIOSCIENCES

Future SMRT systems could reveal interactions.

hospital in Rotterdam, the Netherlands, and lead author of one of the studies.

In his protocol, Poot pulled interacting proteins from cells using nuclear extracts expressing different Flag-tagged transcription factors. He added a nuclease to his reactions to remove DNA and eliminate possible artefacts caused by proteins binding to it. "Transcription factors bind to DNA so you are likely to pull out DNA-binding factors that are not directly interacting," he explains. Purifying many different transcription factors with the same protocol also enabled the researchers to determine which interactions were most likely to be specific. For example, proteins that consistently co-purified with all transcription factors would be treated as unlikely to indicate a genuine interaction.

Calling out false positives — reported interactions that don't actually occur — and false negatives — interactions that do occur but are not picked up by the experimental protocol or are discarded — is one of the main challenges in the field. "Normally when you do a coIP followed by MS you will get hundreds of protein candidates interacting with any one bait," says Wade Harper, a cell biologist at Harvard Medical School in Boston, Massachusetts. "When you weed out all the stochastic and non-specific interactions you end up with many fewer proteins. Some proteins in large complexes might have 30–50 partners, others only 4–5."

One way in which researchers increase the accuracy of their results is to use more than one method (for example, Y2H plus LUMIER) to detect the interactions. But the definition of a 'real' interaction depends on the context. "Does a real interaction mean that two proteins interact if they are placed next to each other in a test tube, or that they must interact in a cell? Or does real mean that the interaction should have a biological function?" asks Ideker. Researchers can home in on functional interactions by combining data on interactions with other types of biological information, such as genetic interactions, protein localizations or gene expression. For instance, proteins whose genes are co-expressed are likely to interact with each other or to be part of the same complex or pathway.

Many tools are available on the web for integrating different types of information about a given protein or gene. One is GeneMANIA, developed by Bader's group in collaboration with Quaid Morris, a computational biologist also at the University of Toronto. A user enters the gene names into GeneMANIA; the program provides a list of genes that are functionally similar or have shared properties, such as similar expression or localization, and then displays a proposed interaction network, showing relationships among the genes and the type of data used to gather that information. The user can click on any node to obtain information about the gene and on any link to obtain information about their relationship (such as citations for any published studies or other sources of data). "It's like a Google for genetic and protein information," says Bader.

Other web-based interfaces that predict gene functions include STRING (http://string-db.org) developed at the European Molecular Biology Laboratory in Heidelberg, Germany. It hunts for protein interactions on the basis of genomic context, high-throughput experiments, co-expression and data from the literature.

KEEPING SCORE

To select real protein–protein interactions, Harper and some members of his lab, Matt Sowa and Eric Bennett, developed a software platform called CompPASS to assign confidence scores to an interaction detected by MS[5]. CompPASS takes data sets of interacting proteins (including those identified in experiments) and measures frequency, abundance and reproducibility of interactions to calculate the score.

This year, Harper used CompPASS to identify interactions among proteins involved in autophagy, the process by which cellular proteins and organelles are engulfed into vesicles and delivered to the lysosome to be degraded. Starting with 32 proteins known to have a role in autophagy, they identified 2,553 interacting proteins using coIP–MS. CompPASS then narrowed the list down to 409 high-confidence interacting proteins with 751 interactions[6].

Ideker's group used a different approach to map interactions among human mitogen-activated protein kinases (MAPKs), which respond to external stimuli and regulate cell function. Having used Y2H to identify more

Tools for the search

HYBRIGENICS

Methods such as the yeast two-hybrid system allow scientists to work out which proteins interact.

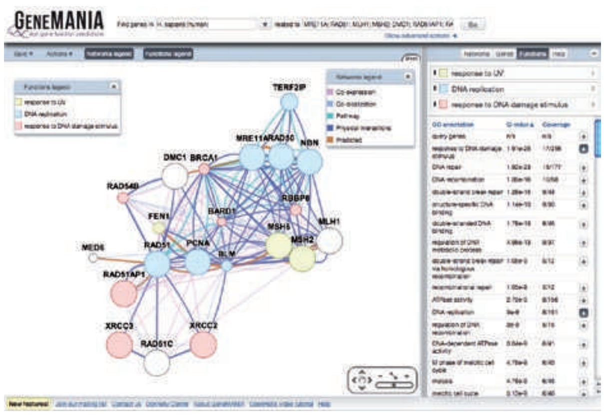
The two main methods for finding protein–protein interactions are the yeast two-hybrid (Y2H) system and co-immunoprecipitation followed by mass spectrometry. Several companies sell reagents for both approaches. Invitrogen of Carlsbad, California, sells the ProQuest Two-Hybrid System with Gateway Technology. This is based on Y2H, with modifications to decrease false-positive results and allow rapid characterization, says the company. Other firms provide vectors used to produce proteins with affinity tags, which can easily be immunoprecipitated along with other interacting proteins. A polypeptide tag called Flag is popular among researchers, and Sigma Aldrich of St Louis, Missouri, provides several Flag-genes for purchase. Promega in Madison, Wisconsin, has the HaloTag technology, in which a protein of interest is expressed in fusion with a tag protein engineered from a bacterial enzyme. This tag can be used to purify the protein, and any interacting with it, by binding to a resin. The tag is cleaved off using a protease.

For researchers who don't have the time or infrastructure to do the experiments, companies such as Hybrigenics in Paris and Dualsystems Biotech of Schlieren, Switzerland, offer Y2H-based screening. "We have complex libraries with ten times more independent clones than most other libraries, which we screen to saturation. And rather than screening full-length proteins, we screen for interactions with domains," says Etienne Formstecher, director of scientific projects and sales at Hybrigenics. "Full-length proteins can have some domains buried and not available to interact, at least in yeast where you may not have signals to unlock a closed protein conformation." A customer is given a list of proteins that interact with the protein of interest; it indicates which domains are making contact and provides a confidence score for each interaction.

Innoprot in Derio, Spain, provides an interaction service using tag-based purification designed for high-throughput analysis. And Invitrogen's ProtoArray Protein–Protein Interaction Service uses microarrays containing more than 9,000 human proteins to identify proteins that interact with any protein of interest. L.B.

lysate are digested. Using tandem MS, researchers can compare the levels of native and synthetic peptides in a cell to obtain a measure of the amount of native proteins present. Synthetic peptides can also be prepared with modifications, such as extra phosphate groups, to measure the number of post-translationally modified proteins. "We are pursuing the dynamics of protein networks by quantifying changes in the amount of proteins present in specific protein complexes," says Harper. "Techniques such as AQUA provide an accurate and sensitive measure of how the stoichiometry of components within complexes that make up a network are altered in response to a stimulus."

The team used the approach to describe the rearrangements that occur in the protein network of cullin-RING ubiquitin ligases, enzymes that regulate protein turnover, under various cellular conditions[11].



Web tools such as GeneMANIA integrate data on a protein or gene.

DEVELOPMENTS IN DIAGNOSTICS

Changes in protein–protein interaction networks may provide information about the mechanisms of disease. Last year, Wrana applied the network approach to the diagnosis of breast cancer. He used microarrays to measure genome-wide protein expression in the tumours of people with breast cancer, and then overlaid the expression data on the network diagram of the human interactome.

Wrana had noted that 'hub' proteins, defined as those that interact with more than four others, can be grouped into two categories depending on whether they are expressed at the same time as the proteins with which they interact. When they looked at breast-cancer samples, Wrana and his colleagues found that certain hub proteins were in a different category in breast-cancer patients with a good prognosis than in those with a poor prognosis.

Thus, by overlaying the expression pattern of a cancer cell from an individual patient onto the human interactome network, Wrana could predict a patient's prognosis. "We found that the detection of global changes in network organization is more predictive of outcome than is gene expression alone," says Wrana. "We have now applied this method to other tumour models and obtained similar results."

KAYAK (kinase activity assay for protein profiling) is another approach to developing diagnostic tools for cancer on the basis of the functional consequences of the interaction between a protein, in this case a kinase, and its substrate. In this method, up to 90 peptide substrates for kinases are used to simultaneously measure the addition of phosphate groups to proteins in a cell lysate — in essence providing a 'phosphorylation signature' for that particular cell. "The readout is so sensitive and so quantitative that even small differences are teased out," says Gygi, who helped to develop the method[12].

According to Gygi, the biggest application of KAYAK might be in tumour classification. "Biopsies or excised tissues can be profiled for kinase activities with pinpoint accuracy. These patterns could contribute towards personalized drug treatments based on dysregulated kinase pathways," he says.

The combination of different types of data and technologies should continue to fill in the empty spaces of the current human interactome map. The picture may never be complete, but it will continue to provide insights into cellular mechanisms of health and disease. "I think that the network we have is dense enough for us to start doing studies to classify disease states," says Wrana. "As the networks become better and coverage improves, the accuracy of diagnosis will also improve." ■

Laura Bonetta *is a freelance science writer based in Garrett Park, Maryland.*

1. Venkatesan, K. *et al. Nature Meth.* **6,** 83–90 (2009).
2. Fields, S. & Song, O.-K. *Nature* **340,** 245–246 (1989).
3. van den Berg, D. L. C. *et al. Cell Stem Cell* **6,** 369–381 (2010).
4. Pardo, M. *et al. Cell Stem Cell* **6,** 382–395 (2010).
5. Sowa, M. E., Bennett, E. J., Gygi, S. P. & Harper, J. W. *Cell* **138,** 389–403 (2009).
6. Behrends, C., Sowa, M. E., Gygi, S. P. & Harper, J. W. *Nature* **466,** 68–76 (2010).
7. Bandyopadhyay, S. *et al. Nature Meth.* **7,** 801–805 (2010).
8. Nomura, D. K., Dix, M. M. & Cravatt, B. F. *Nature Rev. Cancer* **10,** 630–638 (2010).
9. Weerapana, E. *et al. Nature* **468,** 790–795 (2010).
10. Barrios-Rodiles, M. *et al. Science* **307,** 1621–1625 (2005).
11. Bennett, E. J. *et al. Cell* (in the press).
12. Yonghao, Y. *et al. Proc. Natl Acad. Sci. USA* **106,** 11606–11611 (2009).
13. Uemura, S. *et al. Nature* **464,** 1012–1017 (2010).

Real-time analysis

In November, Pacific Biosciences of Menlo Park, California, commercially released its third-generation DNA-sequencing platform, based on its single-molecule, real-time (SMRT) technology. A single DNA polymerase bound to a DNA template is attached to a tiny chamber illuminated by lasers, and nucleotides labelled with coloured fluorophores are introduced to it. As the polymerase incorporates them, each base is held for a few microseconds, while the fluorophore emits coloured light corresponding to the base identity. SMRT technology could also be used to analyse biomolecules other than DNA, and could become a common tool for detecting protein interactions, with some unique features. "This technology can detect relatively weak interactions," says Jonas Korlach, a scientific fellow at Pacific Biosciences, adding that it could pick out interactions that happen so quickly that they can't be identified by current methods.

As a step towards such applications, Joseph Puglisi, a structural biologist at Stanford University School of Medicine in California, and his group, with scientists at Pacific Biosciences, observed transfer RNAs binding to single ribosomes in real time[13]. In an unpublished follow-up, Puglisi's group has used SMRT technology to watch interactions between transfer RNAs, ribosomes and protein factors to determine how the translation machinery synthesizes proteins. "We have just seen the tip of the iceberg in terms of applications," says Korlach. L.B.


PACIFIC BIOSCIENCES

Future SMRT systems could reveal interactions.

ANTIBODY COMPANIES

Company	Products/Activity	Location	URL
AbD Serotec	Sources for research and diagnostic antibodies	Kidlington, UK	www.ab-direct.com
Abgent	Antibodies; customized antibodies; protein-expression services	San Diego, California	www.abgent.com
Active Motif	Antibodies and reagents for transcriptional regulation and epigenetics	Carlsbad, California	www.activemotif.com
AdiMab	Human antibodies produced in yeast	Lebanon, New Hampshire	www.adimab.com
AnaSpec	Off-the-shelf and customized antibodies and peptides	Fremont, California	www.anaspec.com
BioGenes	Customized antibodies; immunoassays; peptides	Berlin, Germany	www.biogenes.de
Charles River	Customized antibodies	Wilmington, Massachusetts	www.criver.com
Epitomics	Customized and off-the-shelf antibodies	Burlingame, California	www.epitomics.com
Fitzgerald Industries	Antibody distributors	Acton, Massachusetts	www.fitzgerald-fii.com
GeneTel	Antibodies for proteomics; chicken-antibody specialists	Madison, Wisconsin	www.genetel-lab.com
GENOVAC	Antibodies against G-protein-coupled receptors; genetic immunization	Freiburg, Germany	www.genovac.com
GenWay Biotech	Primary and secondary antibodies (produced in chickens)	San Diego, California	www.genwaybio.com
GTC Biotherapeutics	Human-antibody production in milk from transgenic animals	Framingham, Massachusetts	www.gtc-bio.com
Harlan	Customized antibodies, peptide synthesis, hybridoma development	Indianapolis, Indiana	www.harlan.com
Innovative Research	Customized and off-the-shelf antibodies	Novi, Michigan	www.innov-research.com
Lonza Biologics	Therapeutic antibodies and recombinant proteins	Basel, Switzerland	www.lonzabiologics.com
Mabtech	Antibodies for ELISA and ELISA Spot applications; cytokines	Nacka Strand, Sweden	www.mabtech.com
Maine Biotechnology Services	Hybridoma, polyclonal and monoclonal antibody development, ascites and *in vitro* production	Portland, Maine	www.mainebiotechnology.com
Millipore	Antibodies; secondary reagents, probes, tests and kits	Billerica, Massachusetts	www.millipore.com ●
MorphoSys	Human antibodies; HuCAL and HuCAL gold antibodies	Martinsried, Germany	www.morphosys.com
New England Peptide	Customized and off-the-shelf antibodies; peptide arrays	Gardner, Massachusetts	www.newenglandpeptide.com
Novus Biologicals	Customized and off-the-shelf antibodies	Littleton, Colorado	www.novusbio.com
OriGene	Proteins and customized antibodies; collaborative antibody production	Rockville, Maryland	www.origene.com ●
ProSci	Off-the-shelf and customized antibodies; immunoassays	Poway, California	www.prosci-inc.com
SDIX	Customized and off-the-shelf antibodies	Newark, Delaware	antibodies.sdix.com
SouthernBiotech	Catalogue and customized antibodies and reagents	Birmingham, Alabama	www.southernbiotech.com

SERVICES

Company	Products/Activity	Location	URL
Biomol	Services for chemical synthesis, cell culture and antibody production	Hamburg, Germany	www.biomol.de
Dualsystems Biotech	Yeast two-hybrid screening service and products	Schlieren, Switzerland	www.dualsystems.com
Hybrigenics	Protein-interaction services based on yeast two-hybrid system	Paris, France	www.hybrigenics.com
Innoprot	Protein-interaction service using affinity-tag-based purification	Bizkaia, Spain	www.innoprot.com
Invitrogen	Protein-interaction services based on proteome arrays	Carlsbad, California	www.invitrogen.com

SOFTWARE ANALYSIS TOOLS

Company	Products/Activity	Location	URL
Ariadne	Pathway analysis	Rockville, Maryland	www.ariadnegenomics.com
Array Genetics	Protein database; genomics, proteomics and microarray analysis	Newtown, Connecticut	www.arraygenetics.com
BIOBASE	Biological databases; analysis tools for gene expression	Wolfenbüttel, Germany	www.biobase-international.com
BioCyc	Pathway analysis software plus collection of databases	Menlo Park, California	www.biocyc.com
BioDiscovery	Microarray analysis and discovery software	El Segundo, California	www.biodiscovery.com
Bitplane	Image-analysis software; filament tracers and auto-aligners	Zurich, Switzerland	www.bitplane.com ●
Bruker Daltonics	Instruments, software and consumables for mass spectrometry	Billerica, Massachusetts	www.bdal.com
Ceiba Solutions	Gene-expression data-analysis system	Cambridge, Massachusetts	www.ceibasolutions.com
Cyberell	Bioinformatics software and services	Helsinki, Finland	www.cyberell.com
Dalicon	Bioinformatics software for managing and analysing large-scale data	Nijmegen, the Netherlands	www.dalicon.com
ePitope Informatics	Epitope predictions; protein-analysis software	Hexham, UK	www.epitope-informatics.com

SOFTWARE ANALYSIS TOOLS

Company	Products/Activity	Location	URL
Genedata	Bioinformatics systems and services for sequence and genome analysis	Basel, Switzerland	www.genedata.com
GeneGO	Data mining and analysis solutions in systems biology	St Joseph, Missouri	www.genego.com
Genomatix	Software for studying molecular mechanisms of gene regulation	Munich, Germany	www.genomatix.de
Geospiza	Bioinformatics software; tools for sequence assembly and analysis	Seattle, Washington	www.geospiza.com
IBM Life Sciences	Database integration; data-management systems	Armonk, New York	www.ibm.com
Ingenuity Systems	Information solutions and custom service for life-sciences research	Redwood City, California	www.ingenuity.com
LGC Genomics	Analysis, annotation and visualization of full chromosome sequences	Berlin, Germany	www.lgcgenomics.com
Life Technologies	Maps of interconnected biological signalling and metabolic pathways	Carlsbad, California	www.lifetechnologies.com
LifeSpan BioSciences	Database for drug-target discovery; antibodies and custom antibodies	Seattle, Washington	www.lsbio.com
Metaphorics	Database of protein–ligand structural information	Aliso Viejo, California	www.metaphorics.com
MiraiBio	Bioinformatics solutions	South San Francisco, California	www.miraibio.com
Parc Research	Software for the analysis of glycan mass spectrometry data sets	Palo Alto, California	www.parc.com
Partek	Pattern recognition and interactive visualization software	St Louis, Missouri	www.partek.com
Premier Biosoft	Sequence analysis, primer design and two-hybrid protein interactions	Palo Alto, California	www.premierbiosoft.com

GENERAL MOLECULAR BIOLOGY REAGENTS

Company	Products/Activity	Location	URL
Agilent Technologies	Reagents and instruments for life-sciences research	Santa Clara, California	www.agilent.com
Applied Biosystems	Reagents for molecular- and cell-biology research	Carlsbad, California	www.appliedbiosystems.com
Attagene	Transcription-factor profiling system; software	Morrisville, North Carolina	www.attagene.com
BD Biosciences	Research reagents, bioimaging systems, instrumentations	Franklin Lakes, New Jersey	www.bd.com
Bio-Rad	Products, instruments and software for life-sciences research	Hercules, California	www.bio-rad.com
BMG Labtech	Microplate and array readers and handling systems	Offenburg, Germany	www.bmglabtech.com
Cambrex	Products for molecular- and cell-biology research	East Rutherford, New Jersey	www.cambrex.com
Cole-Parmer	Variety of instruments and reagents	Vernon Hills, Illinois	www.coleparmer.com
EMD Biosciences	Calbiochem, Novabiochem and Novagen product lines	San Diego, California	www.emdbiosciences.com
Enzo Life Sciences	Consumables and assays for molecular biology, gene expression and genomic analysis	New York, New York	www.enzo.com
Horiba	Spectroscopy systems and accessories	Kyoto, Japan	www.horiba.com
Invitrogen	Reagents for cell and molecular biology	Carlsbad, California	www.invitrogen.com
Irvine Scientific	Defined media for cell-culture applications; custom media services	Santa Ana, California	www.irvinesci.com
Lonza	Molecular-biology reagents and systems; advanced chemical synthesis	Basel, Switzerland	www.lonza.com
Metrohm	Laboratory instruments; consumables	Riverview, Florida	www.metrohmusa.com
Merck	Chemicals, kits and reagents	Darmstadt, Germany	www.merck.de
MP Biomedicals	Reagents and chemicals for research	Solon, Ohio	www.mpbio.com
New England BioLabs	Molecular-biology-related reagents, kits and enzymes	Ipswich, Massachusetts	www.neb.com ●
Pacific Biosciences	Platform for single molecule, real-time detection of biological events	Menlo Park, California	www.pacificbiosciences.com
PerkinElmer	Instruments, reagents and kits for life sciences	Waltham, Massachusetts	las.perkinelmer.com
Princeton Separations	DNA purification columns and reagents, fluorescent protein labelling kits	Adelphia, New Jersey	www.prinsep.com
Promega	Chemicals for mass spectrometry	Madison, Wisconsin	www.promega.com
Sigma-Aldrich	Reagents for chemistry and molecular-biology research, including protein arrays	St Louis, Missouri	www.sigmaaldrich.com ●
Takara Bio	Reagents, kits and consumables for molecular biology	Shiga, Japan	www.takara-bio.com ●
Thermo Scientific	Instruments and reagents for life-sciences research	Waltham, Massachusetts	www.thermo.com
Tocris Bioscience	Chemicals for life-sciences research; contract research services	Bristol, UK	www.tocris.com
USB	Chemicals and reagents for molecular biology	Cleveland, Ohio	www.usbweb.com
Wako Chemicals USA	Speciality chemicals; clinical diagnostic reagents	Richmond, Virginia	www.wakousa.com ●

● see advertisement

CAREERS

TURNING POINT Biophysicist explains how going overseas enriched his career p.859

IRELAND Biomedical institute set to recruit for diagnostic research p.859

NATUREJOBS For the latest career listings and advice www.naturejobs.com

WWW.JUPITERIMAGES.COM

PUBLIC HEALTH

Food-safety sentinels

Disease outbreaks in recent years have revealed the vulnerability of food supplies. But they offer opportunities for those interested in waging war on microbes.

BY LAURA CASSIDAY

In the summer of 2009, a particularly nasty strain of the bacterium *Escherichia coli*, known as O157:H7, infected at least 74 people across 32 US states. They had painful abdominal cramps and diarrhoea, and about half of those affected were ill enough to need hospitalization. Initial attempts to trace the outbreak to a particular food product left epidemiologist Karen Neil perplexed.

"Normally, *E. coli* O157 outbreaks are associated with ground beef or leafy greens like spinach or lettuce," says Neil, who works at the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia. But according to questionnaires filled out by the patients, not everyone had eaten the usual suspects in the days before their illness. So Neil conducted in-depth interviews with five patients, and revealed a common factor: all had eaten the same brand of refrigerated cookie dough. "This was the first time that raw cookie dough was associated with an *E. coli* O157 outbreak," says Neil.

Such detective work is common for those charged with safeguarding the food supply from biological and chemical contaminants. At each step along the chain from farm to table, researchers are needed to quickly detect and respond to problems when they arise. Scientists such as Neil work on the front lines of outbreak response, whereas others focus on prevention or detection strategies or on understanding how food-borne pathogens make people sick. Employers seek expertise in areas ranging from public health and epidemiology to microbiology and behavioural sciences. Although federal funding for food-safety research in the United States has remained mostly flat since a spike after the terrorist attacks on 11 September 2001, the breadth of the field means that there is a consistent demand for researchers in government, academia and industry.

GOOD GROWTH

As food imports from around the globe increase, more personnel will be needed to ensure their safety — and government scientists are often the first line of defence. Inspectors employed by the US Department of Agriculture (USDA), the Food and Drug Administration (FDA) or private companies visit factories, slaughterhouses, restaurants and ports. There they examine sanitary conditions and manufacturing processes, ensure proper food labelling and collect samples for laboratory testing. The USDA, for example, employs more than 7,500 food inspectors nationwide, with entry-level positions requiring a bachelor's degree or one year of job-related experience in the food industry. The FDA's Center for Food Safety and Applied Nutrition (CFSAN) in College Park, Maryland, shares monitoring duties with the USDA. The CDC works with both agencies, as well as with state governments, to investigate outbreaks of infectious disease.

Detecting and identifying food-borne pathogens is a slow process. Inspectors must swab a food sample onto a Petri dish and culture it for at least six hours before laboratory analysis. According to Steven Musser, director of the Office of Regulatory Science at the CFSAN, his office has a high demand for scientists with experience in developing tools to detect food contaminants, especially portable devices that work more quickly. "It's impossible to screen all of the material," he says. "But the more tools we can put in the hands of inspectors in the field, the more products we can look at. We're looking to hire bioinformaticians who know how to handle very large data sets, because right now we can generate data much faster than we ▶

M. FRASER

can analyse them."

"Right now we can generate data much faster than we can analyse them."
Steven Musser

Researchers also need tools to trace outbreaks to their source quickly. In the United States, physicians report suspected food-borne illnesses to state public-health departments, where scientists culture microorganisms from patients' stool samples and analyse pathogen DNA with pulsed-field gel electrophoresis. They report genetic signatures of detected pathogens to a CDC database that links the PulseNet national network of public-health labs and is used to identify clusters of outbreaks that may be linked to a common source.

Closely related strains of some microorganisms, such as *Salmonella*, can't be differentiated using pulsed-field gel electrophoresis. So the CFSAN Office of Regulatory Science employs about 130 scientists who work closely with biotech companies to develop techniques for fingerprinting food-borne pathogens, including whole-genome sequencing, proteomic analysis and mass spectrometry. "Data analysis is one of the big bottlenecks of this work," says Musser.

Also in demand are researchers with expertise in analytical instrument development, food microbiology and mass spectrometry. Musser says that the number of new research positions available in 2011 will depend on the level of FDA funding in the federal budget. "We typically don't start the hiring process until Congress passes the budget in the springtime or summertime," he says.

PUBLIC CONCERN

Food-safety research at the FDA will get a boost if Congress and President Barack Obama approve the Food Safety Modernization Bill, which the US House of Representatives passed in 2009. The Senate passed a similar bill on 30 November 2010. If lawmakers can work out discrepancies, the legislation could be signed into law later this year. The Senate version of the bill calls for increased surveillance and testing of foods by the FDA and includes provisions for at least 4,000 new field staff at the CFSAN and the FDA Center for Veterinary Medicine in Rockville, Maryland, in 2011, with further yearly increases until 2014. The bill also appropriates US$825 million for food-safety work at these centres in the first fiscal year.

In general, funding for food-safety research waxes and wanes with public concern over disease outbreaks. In the 1990s, 'mad cow disease' galvanized food-safety research in Europe. At the same time, US federal funding for food safety increased by more than 60%, spurred by a deadly 1993 outbreak of *E. coli* at the fast-food restaurant chain Jack in the Box. With the terrorist attacks of 11 September 2001, funding priorities shifted to food security. More recently, much-publicized outbreaks of *Salmonella* in peanut products and eggs seem to have refocused public attention on food safety, as reflected by the pending legislation.

But government support is not enough. Collaborations among government, academia and industry are essential to keep food safe. In Europe, scientists at universities in the Netherlands collaborate with industry researchers through the Top Institute of Food and Nutrition based in Wageningen, a public–private partnership funded by the Dutch government that fosters industrially relevant innovations in food quality and safety. In January 2011, the Top Institute will announce vacancies for graduate students and postdocs in 20 new research projects at member institutes, says Marcel Zwietering, a professor of food microbiology at Wageningen University, a participant in the partnership that has an internationally renowned programme in food-safety research.

Basic-science researchers often interact with industry and government. Michael Peck, programme leader at the Institute of Food Research in Norwich, UK, says that much of his time is spent working with industry and food-safety regulators to apply his research findings on the physiology and molecular biology of *Clostridium botulinum*, the food-borne pathogen responsible for botulism. Similarly, researchers at the University of Georgia Center for Food Safety in Griffin collaborate closely with scientists at the nearby CDC. "After my lab developed the first test for detecting *E. coli* O157 in food, the CDC began sending us food samples associated with outbreaks," says Michael Doyle, director of the University of Georgia centre. "Now that some of the state health departments have gotten good at detecting harmful microbes in food, they usually only send us samples in which the organisms are really hard to isolate."

"State health departments have gotten good at detecting harmful microbes in food."
Michael Doyle

Frequent collaborations with other sectors mean that people skills and the ability to work well in a team are valuable assets for those in the food-safety field, notes Neil.

Food-safety researchers in industry have another motivation: company profits. A food recall can cost a company tens of millions of dollars in lost profits and liability lawsuits and even force them out of business, as demonstrated by the 2008–09 *Salmonella* outbreak at the now-defunct Peanut Corporation of America. "Safety will always be an important thing for food companies to guarantee, because one big problem can ruin 100 or more years of existence of a brand," says Zwietering.

DODGING DISASTER

To avoid such scenarios, industry scientists conduct routine testing of products, perform environmental monitoring at factories, and work to develop more efficient detection methodologies. Technicians may not need even a bachelor's degree for entry, whereas team leaders typically have PhDs in areas such as food science, chemistry, microbiology or chemical engineering. "Many of our former students are now in food-safety positions with major food companies," says Shaun Kennedy, director of the National Center for Food Protection and Defense at the University of Minnesota in St Paul. Zwietering notes that in the Netherlands, although the number of students interested in food and agriculture has dropped, the demand from industry remains unchanged.

Food companies typically have more resources than universities for training researchers in specific skills, and provide ample opportunities for career advancement. Scott Hood, senior manger for microbiology and thermal processing at General Mills in Minneapolis, Minnesota, says that at a large global food company such as his, it is common to spend a couple of years in the lab, then maybe a year or two at a plant seeing the daily challenges of the production environment, before coming back to headquarters to work on food safety and quality at the corporate level.

But when every food-industry executive's worst nightmare — a suspected outbreak of food poisoning — does occur, scientists from all sectors work together. The company's food scientists work overtime with government agencies to trace the source of the outbreak to a specific plant and production run.

Like many food-safety researchers, Neil savours the opportunity to make an impact. "It's a very important and exciting field to be in right now," she says. Before joining the staff at the CDC in June 2010, Neil was a postdoctoral fellow in the CDC's Epidemic Intelligence Service, a two-year training programme for health professionals interested in applied epidemiology. She conducted field investigations in Rhode Island, Missouri and Uganda.

"Sometimes you don't know in the morning that you'll be leaving for an outbreak investigation that afternoon," she says, "which makes the job even more interesting." ■

Laura Cassidy *is a freelance writer based in Hudson, Colorado.*

TURNING POINT
Charalampos Kalodimos

C. KALODIMOS

A biophysicist at Rutgers University in Piscataway, New Jersey, Greek-born Charalampos Kalodimos is the first person to win two key young-investigator awards: the Biophysical Society's 2011 Michael and Kate Bárány Award in September, and the Protein Society's award last year.

Are you a natural-born scientist or a convert?
I didn't know I wanted to be a scientist until my first year of graduate studies at the Curie Institute in Paris. There, for my PhD, I began working in bioinorganic chemistry, modelling the binding sites of haemoglobin and myoglobin. I became mesmerized by the notion that, as a scientist, you can be the first person to discover something. But to do so, you must excel at two things: finding interesting questions, and strategically designing approaches to address those questions.

What was your crucial early-career decision?
As a postdoc, I decided to join Robert Kaptein's lab at Utrecht University in the Netherlands. He was one of the pioneers of structural biology, helping to develop the tools we use to determine the three-dimensional structure of biomolecules. Joining his lab was so important because I was introduced to biomolecular nuclear magnetic resonance (NMR). Using NMR, we worked to resolve interesting biological phenomena, such as protein–DNA interactions and DNA regulation. It was an exciting environment because I had the freedom to pursue any kind of question. I learned that what I really like is figuring out the structures controlling how molecular systems work. We tend to use NMR imaging because it is so powerful.

Has your career benefited from tackling several research projects at once?
I am always conducting exploratory research and writing grants to get the next project funded. Sometimes it can take 3–4 years to get enough preliminary data to get the funding. Although some mentors cautioned me that this strategy might be too risky, it worked for me. I think it's just as risky if researchers focus all their resources on getting one project to work. What was the worst thing that could happen? I had a pretty nice back-up plan — I would go back to Greece and live by the sea.

Why the United States for your first tenure-track position?
Although the scientific infrastructure is good in Europe, junior faculty members are not completely independent at most places. You start off with small budgets. What I like here is that you get more freedom early in your career and get a good start-up package. If you fail, there is no one else to blame. If you do great, you get all the praise. It's very competitive, but I like that. Everything is fair. The only thing that matters is if you do good science. This is the place to be at a young faculty level. I travel to Europe a lot, and I advise young researchers to move to the United States to go after their dream.

Are the awards a career turning point?
It's great to have your peers recognize and appreciate the work you have done. What motivates me the most is knowing I was nominated for the Bárány award by Lewis Kay at the University of Toronto, Canada, and Ad Bax at the US National Institutes of Health, who pioneered NMR tools. In that sense, these awards have only pushed me to work harder as we spend the next five to ten years continuing to investigate large protein complexes, which are still quite challenging — their size and complexity make them difficult to purify.

Have you been tempted by any offers to go elsewhere since winning these two awards?
I'm happy at Rutgers, but if there is any chance that I can be more productive elsewhere in the United States or Europe, then I would certainly consider that. Never say never.

Do you thrive on competition?
Absolutely, 100%. As a scientist investigating biological questions, I invest a lot of time and energy in my work. Knowing that someone else might get to the answer first provides extra motivation to keep pushing harder to get there first. Competitiveness is absolutely required in science. I thrive in that environment. ■

INTERVIEW BY VIRGINIA GEWIN

GRANTS
British funding change

Some UK research funders are opting to give larger and longer-term grants to fewer awardees than before. The government-funded Engineering and Physical Sciences Research Council in Swindon is restructuring its grants on the basis of research showing that larger, longer grants result in a higher publication and citation rate, says spokeswoman Victoria McGuire. And the Science and Technology Facilities Council, also in Swindon, has merged two of its grant mechanisms into a single scheme to provide better long-term support, says spokeswoman Julia Maddock. The Wellcome Trust, a charitable foundation based in London, has created two types of larger grants of up to £425,000 (US$661,000) a year for up to seven years, which will have fewer recipients than its traditional schemes.

IRELAND
Biomedical recruitment

The Biomedical Diagnostics Institute (BDI), an academic–industrial partnership based at Dublin City University, has received a five-year, €19-million (US$25-million) grant that will allow it to recruit up to 40 postdocs and PhD students in surface chemistry, photonics and microfluidics. This is the BDI's second round of funding from Science Foundation Ireland (SFI), a government funding body hit hard by budget cuts this year, and it will support research and development of prototype diagnostic devices. BDI director Michael Berndt says that the institute is diversifying its projects and collaborations to secure outside funding because further SFI support is uncertain.

EUROPEAN UNION
Youth fires innovation

More funding and job openings could arise for early-career scientists in the European Union if an EU council's recommendations are taken up. The Competitiveness Council — which reviews EU economic affairs, industry and scientific research — concluded at a 26 November meeting that young researchers help to stimulate innovation and create a science-based culture, and urged the EU to find ways to attract and retain them. It said researcher mobility is important and must be ensured through retention of pension rights and other benefits. In February 2011, EU nation leaders will discuss economic reform.

THE CLEVEREST MAN IN THE WORLD

Problem solved.

BY TONY BALLANTYNE

“Hi, this is Clark Maxwell, the cleverest man in the world. Ten seconds, €10,000. Off you go!”

“Clark! My name’s Bob. My parachute’s broken! What should I do?”

“Hi Bob. Let me see. GPS has you at 20,000 feet over Arizona. That’s pretty high up! Given a terminal velocity of 180 feet per second, you’ve just under two minutes before you hit the ground.”

“I know! What do I do?”

“That’s a tough one! Give me a minute to think …”

“What? No! Don’t hang …”

Too late. Clark checked the volume of space around Bob on his computer and switched to the next call in the queue.

“Hi, this is Clark Maxwell, the cleverest man in the world. Ten seconds, €10,000. Hit me!”

“This is James Sunderland, chief executive of eToys. Clark, we’ve got a spy in the company. Every new product we develop, our competitors get to market weeks before we do.”

“Spies aren’t your only problem then, you must be very inefficient in terms of product manufacture.”

“Oh. What should we do?”

“That’s two questions, James. Just give me a second …”

Clark called up eToys on a second monitor. Keeping one eye on Bob’s rapid descent, he ran a number of searches in quick succession.

“James! You’d have had the answer yourself if you’d taken the trouble to check your network audit trails. The plans are being deliberately downloaded onto games cartridges as part of the background scenery. Your competitors are buying your secrets wholesale. Now for your second question, may I suggest that you make an appointment with my PA to discuss looking at your company from top to bottom.”

“Uh, sure. Thanks, Clark.”

“Don’t mention it. Bob! How’s it going?”

“Still falling, Clark.”

“I see that. Bob, I want you to look down. Do you see the big lake?”

“Yes. Should I aim for it?”

“No! But don’t you find it beautiful? Calming even?”

“No. Should I?”

“Back soon, Bob … Hi, this is Clark Maxwell, the cleverest man in the world. Ten seconds, €10,000. What’s the problem?”

➲ **NATURE.COM**
Follow Futures on Facebook at:
go.nature.com/mtoodm

JACEY

“This is Lewis. Can’t seem to get a girlfriend, Clark.”

“Hmm. That’s because you’re so self-obsessed. Get a hair cut and start paying attention to someone beside yourself.”

“Hey, can you see me?”

“No. Never seen you in my life, Lewis.”

“Then how do you know that’s true? About the haircut and everything?”

“You’ve got €10,000 to spare and you’re using it to ask a stranger how to get a girl. Anyone who thinks that money solves all their problems is probably pretty self-obsessed. Time’s up!”

“But …”

Clark tapped at his keyboard.

“Hi Bob! I can see you now.”

“How?”

“I’ve taken control of the plane you jumped from.”

“Can you do that?”

“Did I mention I was the cleverest man in the world? Hold it, Bob, I’ll be back in a minute!”

“I don’t have a minute!”

“Hi, this is Clark Maxwell, the cleverest man in the world. Ten seconds, €10,000. How can I be of service?”

“Clark, this is your wife, the smartest woman in the world. Have you walked off with my car keys again?”

“Sorry, Lois. Will you be home tonight?”

“Assuming I get the supercollider fixed. I think I know what’s causing the problem. It’s not its future self it’s interfering with, it’s its past self.”

“Sounds cool, dear. Got to go! Hi, this is Clark Maxwell, the cleverest man in the world. Ten seconds, €10,000. Hit it!”

“Clark, this is Tessa Walkiewicz, Acronym News. We’re doing a report on the acceleration of change and we’d like a few words …”

“Certainly, Tessa. Just a moment … Bob, you’re falling too fast. Hold your arms and legs wide. I’m sure you’ve seen people do it in films!”

“It looks easier in films, Clark.”

“I know! Just do your best! Marianne is jumping out of the plane, right now. She’s got a spare chute for you.”

“What plane?”

“Your plane, Bob. The one you jumped out of. It’s right behind you!”

“Oh! That’s clever!”

“That’s my job … back in a moment, Bob. Tessa! What’s the question?”

“Well, Clark. Given the growth of the Internet and the new paradigms of interconnectivity, people such as yourself are emerging as a powerful force for social change. Plugged into the world’s data streams, you have a view of everything changing from minute to minute.”

“That’s not a question, Tessa.”

“No, that’s an intro, Clark. The question is this: given that people are using services such as yours more and more, does that mean they are getting less intelligent?”

“I hardly think that many people are using my service, Tessa. Not at the prices I charge!”

“Maybe not yours, Clark, but given that the answer to any problem you have is only a phone call away, why should people think for themselves anymore?”

“Let me turn that around, Tessa. When they stop thinking, they stop being people. Got to go!”

“But …”

“Hi, this is Clark Maxwell, the cleverest man in the world. Ten seconds, €10,000. I’m listening!”

“Uh, Clark, this is Marianne. I jumped out of the plane, I’ve attached myself to Bob.”

“Well done Marianne! What’s the problem?”

“It’s my ’chute. It’s failed to open, too. The ground’s looking awfully close.”

“Marianne, thank you! I do like a challenge! Now, listen to me carefully …” ■

Tony Ballantyne *has had short stories published in magazines and anthologies around the world. His latest novel,* Blood And Iron*, is published by Tor UK.*

Global systematics of arc volcano position

ARISING FROM Grove, T. *et al. Nature* **459,** 694–697 (2009)

Global systematics in the location of volcanic arcs above subduction zones[1,2] are widely considered to be a clue to the melting processes that occur at depth, and the locations of the arcs have often been explained in terms of the release of hydrous fluids near the top of the subducting slab (see, for example, refs 3–6). Grove *et al.*[7] conclude that arc volcano location is controlled by melting in the mantle at temperatures above the water-saturated upper-mantle solidus and below the upper limit of stability of the mineral chlorite and in particular, that the arc fronts lie directly above the shallowest point of such melt regions in the mantle. Here we show that this conclusion is incorrect because the calculated arc locations of Grove *et al.*[7] are in error owing to the inadequate spatial resolution of their numerical models, and because the agreement that they find between predicted and observed systematics arises from a spurious correlation between calculated arc location and slab dip. A more informative conclusion to draw from their experiments is that the limits of chlorite stability (figure 1b of ref. 7) cannot explain the global systematics in the depth to the slab beneath the sharply localized arc fronts.

Grove *et al.*[7] hypothesize that arc volcano location is controlled by melting in the mantle at pressure and temperature conditions defined as 'P, T_{melt}' in their figure 1b. Grove *et al.*[7] then use numerical models of subduction zones to predict arc location and its global systematics. They conclude that the agreement between their calculated systematics of arc location and observations of real subduction zones[2,8] validates their hypothesis (figure 3 of ref. 7) but closer inspection of the shape of the P, T_{melt} region casts doubt upon this conclusion. A characteristic feature of subduction-zone models[9] is the narrow thermal boundary layer, sub-parallel to and just above the slab surface, which contains the temperature range of P, T_{melt} (~800–850 °C). For all but the slowest convergence rates, this boundary layer begins close to the depth at which the slab is viscously coupled to the wedge. Hence we should expect the region enclosing P, T_{melt} to be a very thin, continuous layer above the slab, with its shallowest extent at an almost constant depth. The results of Grove *et al.*[7] (green squares in their figure 2) are inconsistent with this expectation, and raise the suspicion of an error in their calculations.

To locate their region of P, T_{melt}, Grove *et al.*[7] determined which nodes of their 2.3 × 2.3-km computational mesh lay within that P–T range. Because those conditions occur within a boundary layer only a few kilometres thick that is inclined at an angle to the mesh, this procedure did not resolve the full extent of the P, T_{melt} region. To check their results, we calculated the temperature fields for subduction zones on a 1 × 1-km grid, then resampled it to both 2.3-km resolution and to 0.25-km resolution. This was done for a range of subduction parameters and for each calculation we determined the P, T_{melt} region and its shallowest point. We found that at 2.3-km resolution, the minimum depth of P, T_{melt} ranged between about 57 and 76 km, consistent with the range found by Grove *et al.*[7]. On the 0.25 × 0.25-km grid, however, the minimum depth was confined between 57 and 61 km (Fig. 1a), consistent with the expectations we describe in the preceding paragraph. At either resolution, the minimum depth of P, T_{melt} is independent of the slab dip and of the convergence rate.

Grove *et al.*[7] compare their calculations with seismic studies, which show that the depth of the slab beneath arcs varies between ~80 and ~150 km (refs 2, 8) and has a negative correlation with the descent speed of the slab (Fig. 1b). The depth to the top of the slab predicted by the hypothesis of Grove *et al.*[7] applied under our recalculations is ~60–75 km, independent of dip or convergence rate (Fig. 1b), and thus does not agree with the observations.

The agreement between model and observations in Grove *et al.*[7] is spurious, and is the result of their choice of variables. Figure 1c recreates their figure 3, which shows the apparent consistency between model and observations, using our recalculated location of arcs. The sine of slab dip is plotted on the x axis, and on the y axis is the arc–trench distance, which for all points (calculated and observed; see

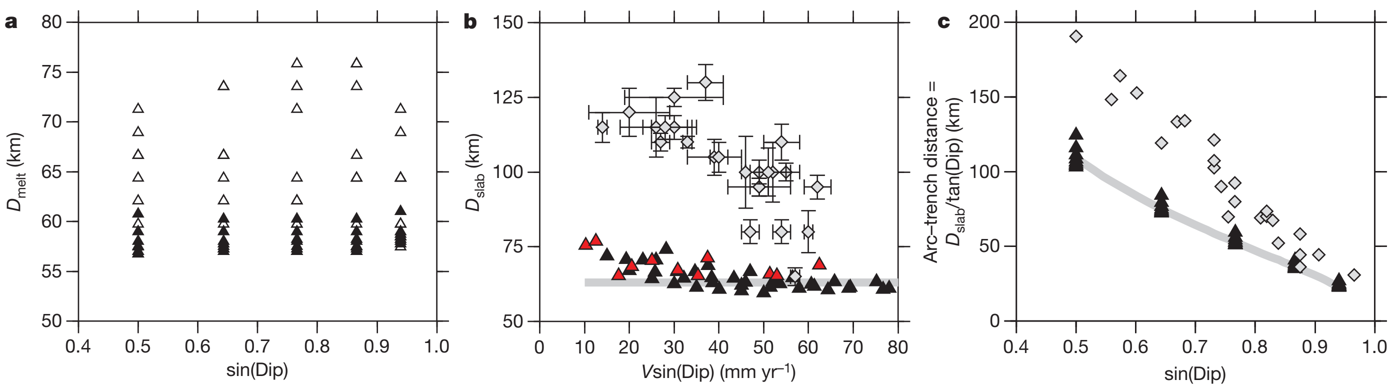


Figure 1 | Arc position versus subduction parameters for data and models. **a**, Calculated depth D_{melt} of the shallowest portion of the P, T_{melt}-based melting field (compare figures 1 and 2 of ref. 7). Calculations were carried out on a 1-km finite-volume mesh[9], for dip of 30° to 70° in steps of 10°, and for convergence rate V from 30 to 100 mm yr^{-1}, in steps of 10 mm yr^{-1}; these ranges include the parameters of the calculations of ref. 7. The points correspond to the minimum depths of melting calculated according to the hypothesis and methods of Grove *et al.*[7] for a 2.3 × 2.3-km resampled grid (open triangles) and for a 0.25 × 0.25-km resampled grid (filled triangles). **b**, Diamonds show the depth of the slab D_{slab}, determined seismologically[2] (error bars as described by ref. 2); filled triangles show the calculated D_{slab} below the locus of shallowest melting, for the 0.25 × 0.25-km resampled grid from panel **a**. The red triangles correspond to the corrected values of D_{slab} for the combinations of dip and convergence rate used by ref. 7 (T. Grove *et al.*, personal communication). The grey line corresponds to a constant D_{slab} = 62 km. **c**, This panel corresponds to the lower 200 km of figure 3 in ref. 7. Points as for panel **b**, plotted for the horizontal distance between the trench and the arc, which is equal to D_{slab}/tan(Dip), the quantity on the y axis of figure 3 of ref. 7. The grey line corresponds to D_{slab} = 62 km and demonstrates the spurious correlation referred to in the main text.

table 1 in ref. 7) is taken as the depth of the slab divided by the tangent of the dip. The presence of the sine of the dip on each axis ensures a spurious correlation; this is illustrated clearly in Fig. 1c by the grey line that corresponds to a constant value of the depth of the slab, $D_{slab} = 62$ km.

Therefore there is no significance in the match between models and observations reported by Grove *et al.*[7], and their conclusion that "the kinematic control on the location of mantle melting is primarily slab dip" (page 696 of ref. 7) is mistaken. Instead, we conclude from their experiments that the limits of chlorite stability (figure 1b of ref. 7) cannot explain the global systematics in the depth of the slab beneath sharply localized arc fronts, which is true for any strongly temperature-dependent process that takes place near the top of the slab, as we have discussed. In ref. 10 we suggest a process that can account for the global systematics in location of the arcs.

Philip C. England[1] & Richard F. Katz[1]
[1]Department of Earth Sciences, Parks Road, Oxford OX1 3PR, UK.
e-mail: philip.england@earth.ox.ac.uk

Received 7 September 2009; accepted 12 April 2010.

1. Tovish, A. & Schubert, G. Island arc curvature, velocity of convergence and angle of subduction. *Geophys. Res. Lett.* **5,** 329–332 (1978).
2. England, P., Engdahl, R. & Thatcher, W. Systematic variation in the depths of slabs beneath arc volcanoes. *Geophys. J. Int.* **156,** 377–408 (2004).
3. Gill, J. *Orogenic Andesites and Plate Tectonics* (Springer, 1981).
4. Tatsumi, Y. & Eggins, S. *Subduction Zone Magmatism* (Blackwell Science, 1995).
5. Iwamori, H. Transportation of H_2O and melting in subduction zones. *Earth Planet. Sci. Lett.* **160,** 65–80 (1998).
6. Tatsumi, Y. The subduction factory: how it operates in the evolving Earth. *GSA Today* **15,** 4–10 (2005).
7. Grove, T., Till, C., Lev, E., Chatterjee, N. & Medard, E. Kinematic variables and water transport control the formation and location of arc volcanoes. *Nature* **459,** 694–697 (2009); erratum **460,** 1044 (2009).
8. Syracuse, E. & Abers, G. Global compilation of variations in slab depth beneath arc volcanoes and implications. *Geochem. Geophys. Geosyst.* **7,** Q05017, doi:10.1029/2005GC001045 (2006).
9. van Keken, P. *et al.* A community benchmark for subduction zone modeling. *Phys. Earth Planet. Inter.* **171,** 187–197 (2008).
10. England, P. C. & Katz, R. F. Melting above the anhydrous solidus controls the location of volcanic arcs. *Nature* **467,** 700–703 (2010).

Competing financial interests: declared none.
doi:10.1038/nature09154

Grove *et al.* reply

REPLYING TO England, P. C. & Katz, R. F. *Nature* **468,** doi:10.1038/nature09154 (2010)

In their Comment England and Katz[1] suggest that our model[2] contains two flaws and that there are additional problems in our thermal models. This Reply points out an important part of our model that England and Katz[1] appear to have missed, addresses their suggestion that there are flaws and discusses whether our thermal models are in error.

The Comment[1] states that we "conclude that the arc fronts lie directly above the shallowest point [that satisfies the P, T_{melt} criterion] in the mantle". This corresponds to our path A in figure 1 of ref. 2. The P, T_{melt} criterion described in ref. 1 refers to the melting that initiates just above the slab over a range of depths, illustrated between paths A and C in figure 1 of ref. 2. As we discussed[2], when these initial melts ascend into the overlying mantle wedge, not all of them will experience a pressure–temperature path that allows them to erupt from an arc volcano on the Earth's surface. The melts formed at the shallowest depths (path A) will encounter cooler mantle as they ascend into the overlying mantle wedge and these melts will freeze. Only melts that ascend into the hottest interior portion of the mantle wedge (such as path B in figure 1 of ref. 2) will undergo sufficient melting to produce arc front volcanoes. To summarize our findings[2], there are two important factors that control the location of arc volcanoes: (1) chlorite dehydration releases H_2O near the slab–wedge interface, and the H_2O ascends into overlying mantle that is above the H_2O-saturated mantle solidus (P, T_{melt} in figure 1b of ref. 2) and (2) the temperature of the overlying mantle wedge increases with decreasing pressure to allow flux melting to continue to high extents and allow these high-extent melts to erupt at arc volcanoes (path B in figure 1 of ref. 2).

England and Katz also state that the agreement that we[2] "find between predicted and observed systematics arises from a spurious correlation between calculated arc location and slab dip" (ref. 1). They attribute this purported spurious correlation in our figure 3 (ref. 2) to the presence of the tangent function on the vertical axis and a sine function on the horizontal axis. Although there is trigonometry involved in the correlation shown on this figure[2], the relations are not spurious and are meaningful. The salient point in our figure 3 is that the beginning of H_2O-saturated melting in our modelling (path A in figure 1a of ref. 2) consistently occurs at a depth of 60–70 km near the slab–wedge interface and is independent of the convergence rate and dip. We point out that these shallowest melts do not reach the surface (figure 1 of ref. 2), nor do they influence the location of volcanoes. Instead, the maximum amount of melting and hence the location of arc volcanoes are controlled by the position of the hottest part of the wedge above a slab. This is the region between paths B and C (figure 1 of ref. 2), the region of maximum melting from our models. The arc–trench distance for paths B to C, and thus the location of arc volcanoes, is close to the values reported by England *et al.*[3] and parallel to the trend of Syracuse and Abers[4]. The distance a given isotherm is from the trench decreases with increasing convergence rate and spans a range of values that are represented in the data of ref. 3. An interesting outcome of our thermal modelling (figure 2 of ref. 2) is that at steep dip angles, paths A and B occur at very similar distances from the trench.

England and Katz say that our thermal modelling results (in figure 2 of ref. 2) "raise the suspicion of an error in [our] calculations" (ref. 1). England and Katz continue with a discussion of grid size in the numerical calculations that they performed, but it is impossible for us or for any reader of the Comment[1] to assess the veracity of their claim that we[2] "did not resolve the full extent of the P, T_{melt} region" (ref. 1). We have verified our modelling methods using the community benchmarks developed for subduction zone modelling[5] and we also find that our model results for the temperature structure near the slab–wedge interface are comparable to those of others who have benchmarked their models, such as Wada and Wang[6], who explicitly considered the issues associated with slab–mantle viscous coupling.

Thus, we disagree with the conclusion reached by England and Katz[1] that "the limits of chlorite stability cannot explain the global systematics in the depth of the slab beneath sharply localized arc fronts". The conclusions we reached in ref. 2 rely on the interplay

of two important controls on hydrous melting in the mantle wedge above subducted slabs: the dehydration of chlorite near the base of the wedge and the temperature structure of the overlying mantle wedge.

T. L. Grove[1], C. B. Till[1], E. Lev[2], N. Chatterjee[1] & E. Médard[3]
[1]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
e-mail: tlgrove@mit.edu
[2]Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA.
[3]Laboratoire Magmas et Volcans, Université Blaise Pascal, Clermont-Ferrand, F-63038, France.

1. England, P. C. & Katz, R. F. Global systematics of arc volcano position. **468,** doi:10.1038/nature09154 (this issue).
2. Grove, T., Till, C., Lev, E., Chatterjee, N. & Medard, E. Kinematic variables and water transport control the formation and location of arc volcanoes. *Nature* **459,** 694–697 (2009); erratum **460,** 1044 (2009).
3. England, P., Engdahl, R. & Thatcher, W. Systematic variation in the depths of slabs beneath arc volcanoes. *Geophys. J. Int.* **156,** 377–408 (2004).
4. Syracuse, E. M. & Abers, G. A. Global compilation of variations in slab depth beneath arc volcanoes and implications. *Geochem. Geophys. Geosyst.* **7,** Q05017, doi: 10.1029/2005GC001045 (2006).
5. van Keken, P. E, *et al.* A community benchmark for subduction zone modeling. *Phys. Earth Planet. Inter.* **171,** 187–197 (2008).
6. Wada, I. & Wang, K. Common depth of slab-mantle decoupling: reconciling diversity and uniformity of subduction zones. *Geochem. Geophys. Geosyst.* **10,** Q10009, doi: 10.1029/2009GC002570 (2009).

doi:10.1038/nature09155

doi:10.1038/nature09565

Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'

Alvin Tamsir[1], Jeffrey J. Tabor[2] & Christopher A. Voigt[2]

Computation underlies the organization of cells into higher-order structures, for example during development or the spatial association of bacteria in a biofilm[1–3]. Each cell performs a simple computational operation, but when combined with cell–cell communication, intricate patterns emerge. Here we study this process by combining a simple genetic circuit with quorum sensing to produce more complex computations in space. We construct a simple NOR logic gate in *Escherichia coli* by arranging two tandem promoters that function as inputs to drive the transcription of a repressor. The repressor inactivates a promoter that serves as the output. Individual colonies of *E. coli* carry the same NOR gate, but the inputs and outputs are wired to different orthogonal quorum-sensing 'sender' and 'receiver' devices[4,5]. The quorum molecules form the wires between gates. By arranging the colonies in different spatial configurations, all possible two-input gates are produced, including the difficult XOR and EQUALS functions. The response is strong and robust, with 5- to >300-fold changes between the 'on' and 'off' states. This work helps elucidate the design rules by which simple logic can be harnessed to produce diverse and complex calculations by rewiring communication between cells.

Boolean logic gates integrate multiple digital inputs into a digital output. Electronic integrated circuits consist of many layered gates. In cells, regulatory networks encode logic operations that integrate environmental and cellular signals[6–8]. Synthetic genetic logic gates have been constructed, including those that perform AND, OR and NOT functions[9–12], and have been used in pharmaceutical and biotechnological applications[13,14]. Multiple gates can be layered to build more complex programs[15–17], but it remains difficult to predict how a combination of circuits will behave on the basis of the functions of the individuals[11,18]. Here we have compartmentalized a simple logic gate into separate *E. coli* strains and use quorum signalling to allow communication between the strains[5]. Compartmentalizing the circuit produces more reliable computation by population-averaging the response. In addition, a program can be built from a smaller number of orthogonal parts (for example transcription factors) by re-using them in multiple cells.

NOR and NAND gates are unique because they are functionally complete. That is, any computational operation can be implemented by layering either of these gates alone[19]. Of these, the NOR gate is the simplest to implement using existing genetic parts. A NOR gate is 'on' only when both inputs are 'off' (Fig. 1a). We designed a simple NOR gate by adding a second input promoter to a NOT gate[20]. Tandem promoters with the same orientation drive the expression of a transcriptional repressor (Fig. 1b). Tandem promoters are common in prokaryotic genomes[21]. This is expected to produce an OR function; however, interference between the promoters can occur (Supplementary Figure 3). The repressor turns off a downstream promoter, which serves as the output of the gate. Both the inputs and the output of this gate are promoters; thus, multiple gates could be layered to produce more complex operations.

Each logic gate is encoded in separate strains of *E. coli*. Acyl homoserine lactone (AHL) cell–cell communication devices are used as signal-carrying 'wires' to connect the logic gates encoded in different strains[4,5,22]. Gates are connected in series where the output of the first gate is the expression of the AHL synthase (*Pseudomonas aeruginosa* PAO1 LasI or RhlI). AHL diffuses through the cell membrane and binds to its cognate transcription factor (*P. aeruginosa* PAO1 LasR or

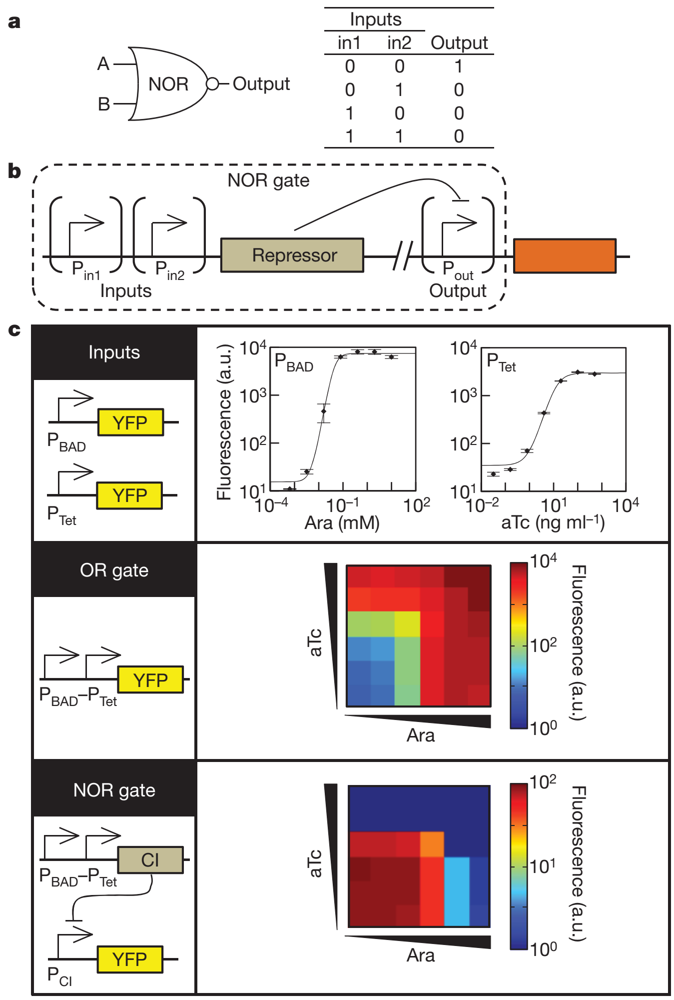


Figure 1 | The genetic NOR gate. **a**, **b**, Symbol, truth table (**a**) and genetic diagram (**b**) of the NOR gate. **c**, The transfer function is defined as the output as a function of input at steady state. The transfer functions of P_{BAD} and P_{Tet} (top), the P_{BAD}–P_{Tet} tandem promoter (middle), and the NOR gate (bottom) are shown. The inducer concentrations for the tandem promoter and NOR gate characterizations are 0, 0.0005, 0.005, 0.05, 0.5 and 5 mM Ara (squares from left to right) and 0, 0.025, 0.25, 2.5, 25 and 250 ng ml^{-1} aTc (squares from bottom to top). Fluorescence values and their error bars are calculated as mean ± s.d. from three experiments. a.u., arbitrary units.

[1]Department of Biochemistry and Biophysics, University of California, San Francisco, California 94158, USA. [2]Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94158, USA.

RhlR). The promoter that is turned on by the transcription factor is used as the input to the next logic gate. These systems have been used previously to program cell–cell communication and have been shown to have little cross-talk[4]. Analogous to a series of electrical gates arrayed on a circuit board, compartmentalization of genetic gates in individual cells allows them to be added, removed or replaced simply by changing the spatial arrangement of the *E. coli* strains.

The stepwise construction of a NOR gate with P_{BAD} and P_{Tet} as the input promoters and yellow fluorescent protein (YFP) as the output gene is shown in Fig. 1c. P_{BAD} and P_{Tet} are activated in the presence of arabinose (Ara) and anhydrotetracycline (aTc), respectively. The individual transfer functions of P_{BAD} and P_{Tet} are measured using flow cytometry (Fig. 1c). An OR gate is constructed by placing the P_{BAD} and P_{Tet} promoters in tandem. P_{BAD}–P_{Tet} demonstrates OR logic with 7,000-fold induction between the 'off' state (−Ara, −aTc) and the 'on' state (+Ara, +aTc). Finally, to convert the OR gate into a NOR gate, the CI-repressor gene is placed under the control of P_{BAD}–P_{Tet} and YFP is expressed from a second plasmid under the control of the CI-repressible P_R promoter. Whereas the OR gates have some characteristics of fuzzy logic, the NOR gates are nearly digital (Fig. 1c).

These OR and NOR gates use promoters as inputs. This feature imparts modularity to the gates; in other words, they can be engineered to respond to different inputs by replacing the promoters. To investigate this, we swapped the input promoters of the logic gates. Figure 2 shows the characterization data for three different tandem promoters: P_{BAD}–P_{Tet}, P_{BAD}–P_{Las} and P_{Tet}–P_{Las}. The promoter P_{Las} is activated by the quorum signal 3OC12-HSL[22] (*N*-3-oxo-dodecanoyl-homoserine lactone). These gates perform as the additive combination of the individual transfer functions of the two input promoters and the CI-repressor NOT gate. The predicted transfer function for the six logic gates shown in Fig. 2 matched the experimental results. One tandem promoter, P_{Tet}–P_{BAD}, did not function as predicted (Supplementary Figure 3). The failure observed for P_{Tet}–P_{BAD} probably arises from some position-dependent interference. This could be the result of the effects of DNA looping, the occlusion of transcription-factor-binding sites or changes in the ratio or stability of output messenger RNAs, among other effects.

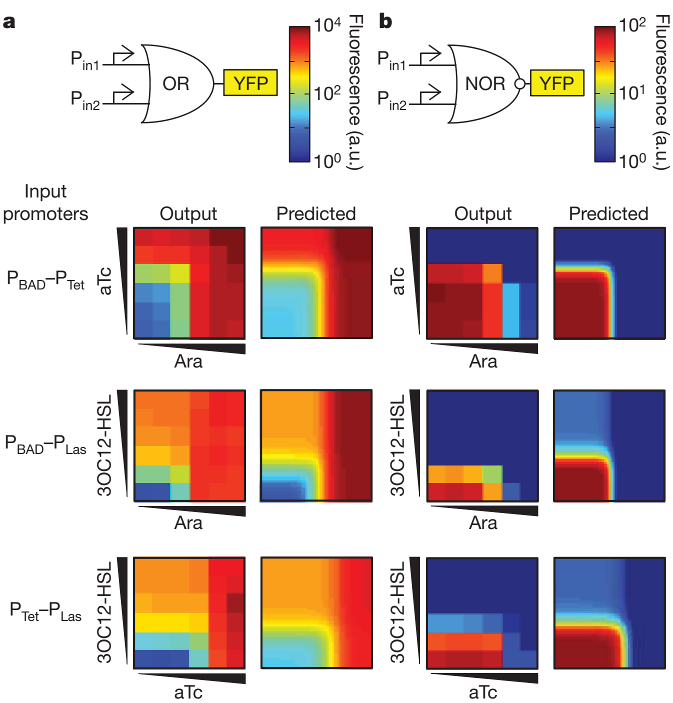


Figure 2 | Input modularity of the gates. **a**, Transfer functions for three OR gates (left) are compared with the predicted transfer function (right). The predicted transfer function is the simple sum of the transfer functions measured for the individual promoters (Supplementary Information). The Ara and aTc concentrations used are the same as in Fig. 1 and those for 3OC12-HSL are 0, 0.001, 0.01, 0.1, 1 and 10 μM (squares from bottom to top). **b**, Transfer functions for three NOR gates (left) are compared with the predicted transfer functions (right). The data represent means calculated from three experiments.

Complex logic can be designed using layers of simpler gates. An XOR gate is built with three NOR gates and a buffer gate (Fig. 3a). The output of an XOR gate is 'on' only when either (but not both) inputs are 'on'. Four strains, each carrying a different logic gate, are used to construct an XOR circuit. The strains are spotted onto an agar plate in the spatial arrangement required to perform this function (Fig. 3b). Cell 1 carries a NOR gate that uses Ara and aTc as inputs and expresses LasI as the output. This allows cell 1 to be wired to the NOR gates in cells 2 and 3 by means of 3OC12-HSL. Cells 2 and 3 use Ara and aTc as their second inputs, respectively. Similarly, the output of the NOR gates in cells 2 and 3 is RhlI, which produces C4-HSL[22] (*N*-butyryl-homoserine lactone). Cell 4 acts as a buffer gate and integrates the outputs from cells 2 and 3 by responding to C4-HSL. The output of a buffer gate is 'on' only when the input is 'on'. The complete circuit consisting of all four strains behaves as a digital XOR gate with respect to the two inputs (Ara and aTc; Fig. 3c, d). Each intermediate colony performs its digital logical operations appropriately, as tested by replacing each output gene with YFP (Fig. 3c).

We constructed a small library of strains that act as simple logic gates, most of which are components of the XOR gate (Fig. 4a). Circuit diagrams showing how all of the sixteen possible two-input logic gates can be constructed using the library are shown in Fig. 4b. Each circuit diagram is reproduced by the spatial arrangement of the component strains. None of these circuits required additional genetic manipulation. The range of induction varies from 5-fold (XOR) to 335-fold (B gate). The dominant contribution to the dynamic range of the complete circuit is due to the intrinsic range of the final circuit (Supplementary Figure 7). For example, the XOR and NAND gates are limited by the output of P_{Rhl}. The addition of a NOT gate to this promoter increases the dynamic ranges of the EQUAL, AND, A IMPLY B and B IMPLY A gates, which is an effect described previously[23]. No degradation in the signal is observed as a function of the number of layers.

The calculations are robust with respect to the distance between colonies and the time and density at which they are spotted (Supplementary Figure 10). This robustness is partially due to the population averaging that occurs, which reduces the effect of cell–cell variation. Despite the variability in the circuit response within a colony, this variability is effectively averaged and thus is not propagated to the next layer of the circuit. The use of chemical signals and population averaging could represent a common design rule for achieving computational operations robust enough to overcome the stochastic limitations of layered circuits in individual cells[24,25]. Another source of robustness is the external clock that is implemented by delaying the spotting of colonies for each layer. Genetic computing is asynchronous and this may result in hazards, that is, transient incorrect outputs that occur as a result of mismatched delays in the circuit[26]. This is apparent when circuits are measured in liquid culture, where the calculation is less robust with respect to timing and cell density (Supplementary Figure 12). To perform the calculation properly, all of the cells need to start in the 'off' state. As layered computation becomes more critical to the design of genetic programs, this will either require the implementation of a genetic clock[27] or the design of programs that are robust to asynchronous computation[28].

Cellular automata have been used to show how simple logic yields complex patterns in the organization of cells[1]. These have been used to model biological pattern formation, development and complex collective behaviour[3,29]. Here we demonstrate that a library of simple gates can be used to form more complex computational operations by linking the gates using diffusible chemical signals. The motif of multiple promoters in tandem driving the expression of a repressor is common

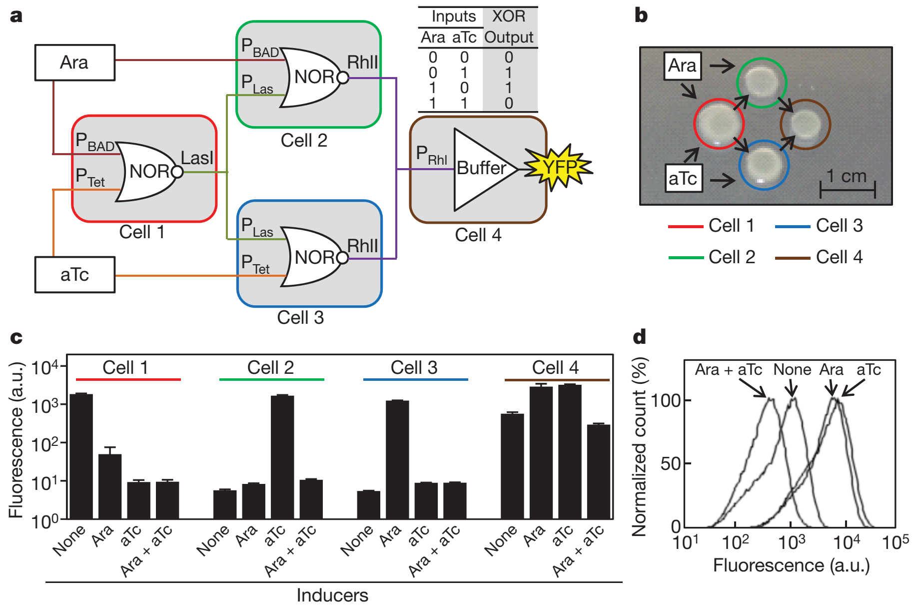


Figure 3 | Construction of an XOR gate by programming communication between colonies on a plate. a, Four colonies—each composed of a strain containing a single gate—are arranged such that the computation progresses from left to right, with the result of each layer communicated by means of quorum signals. The inputs (Ara and aTc) are added uniformly to the plate. **b**, Spatial arrangement of the colonies. **c**, Each colony responds appropriately to the combinations of input signals. Fluorescence values and their error bars are calculated as mean ± s.d. from three experiments. **d**, Cytometry data for the XOR gate (cell 4).

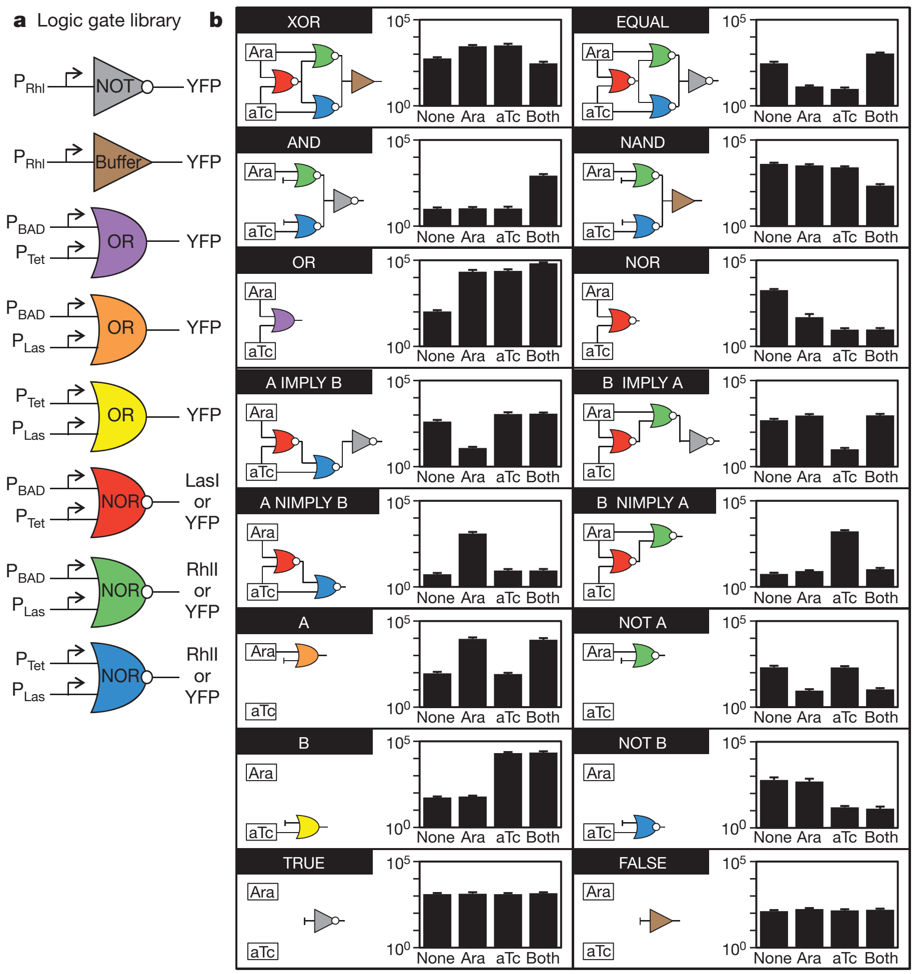


Figure 4 | Construction of all 16 two-input Boolean logic gates. a, Library of simple logic gates carried by different strains (corresponding to plasmids in Supplementary Table 5). **b**, Colonies containing different gates were spotted to mimic the spatial arrangement of each logic circuit (Fig. 3b). For each circuit, the final colony was assayed by flow cytometry for all combinations of inducers added to the plate. The data correspond to the cytometry distributions in Supplementary Figure 6. Fluorescence values and their error bars are calculated as mean ± s.d. from three experiments. NIMPLY, NOT IMPLY.

in genomes[21], and the resulting NOR gates may represent a ubiquitous fundamental unit of biological computation. Although our current ability to create logic gates within a single cell is limited, it may ultimately be possible to encode more complex circuits in individual cells that are then linked by cell–cell communication, akin to logic blocks in field-programmable gate arrays[30]. Together, these principles can be used in the engineering of biological systems to create increasingly complex functions.

METHODS SUMMARY

Strains, plasmids and media. All studies were performed using *E. coli* strain DH10B. Luria–Bertani (LB)-Miller medium (Difco 244610) was used for the assays. The antibiotics used were 50 μg ml^{-1} chloramphenicol (Acros 227920250) and/or 50 μg ml^{-1} kanamycin (Fisher BP906-5). The inducers used were arabinose (Sigma A3256), anhydrotetracycline (Fluka 37919) and 3OC12-HSL (Sigma O9139).

Transfer function characterization. Cells harbouring the appropriate plasmids were incubated in 3 ml of LB broth medium (37 °C, 250 r.p.m. shaking) in culture tubes without the presence of inducers for 18 h. The cultures were then diluted 200-fold into 200 μl fresh LB broth medium (supplemented with appropriate inducers) in a 96-well plate format and incubated for additional 14 h before finally being diluted 100-fold into PBS solution for cytometry analysis.

Plate assay of circuit function. The plate medium was prepared by pouring 12 ml of LB broth agar medium (1.5% agar (Difco 214030), 2.5% LB-Miller) supplemented with inducers (2 mM Ara and/or 500 ng ml^{-1} aTc) into a 100-mm Petri dish (Fisher 08-757-13). Bacterial logic gates were 'fabricated' on the plate by spotting 1 μl overnight culture of appropriate bacterial strains (Supplementary Table 5) to mimic the spatial arrangement of each logic circuit. The distance between each two colonies was set at 7 mm in square grids. Spotting was done with 12-h delay from the previous layer's spotting to ensure communication signals had propagated sufficiently. After 12 h from the last layer's spotting, the whole output colony of the circuit was scraped using inoculating loops and diluted into 10 ml PBS solution for cytometry analysis.

Flow cytometry. All data contained at least 50,000 events, obtained using BD-FACS LSR2. Events were gated by forward and side scatter using MATLAB software. The geometric means of the fluorescence distributions were calculated. The autofluorescence value of *E. coli* DH10B cells harbouring no plasmid was subtracted from these values to give the fluorescence values reported in this study.

Received 2 March; accepted 11 October 2010.
Published online 8 December 2010.

1. Neumann, J. V. *The General and Logical Theory of Automata* (Wiley, 1951).
2. Turing, A. M. The chemical basis of morphogenesis. 1953. *Bull. Math. Biol.* **52,** 119–152 (discussion), 153–197 (1990).
3. Wolfram, S. *A New Kind of Science* 23–113 (Wolfram Media, 2002).
4. Brenner, K., Karig, D. K., Weiss, R. & Arnold, F. H. Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium. *Proc. Natl Acad. Sci. USA* **104,** 17300–17304 (2007).
5. Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434,** 1130–1134 (2005).
6. Li, F., Long, T., Lu, Y., Ouyang, Q. & Tang, C. The yeast cell-cycle network is robustly designed. *Proc. Natl Acad. Sci. USA* **101,** 4781–4786 (2004).
7. Niklas, K. J. The bio-logic and machinery of plant morphogenesis. *Am. J. Bot.* **90,** 515–525 (2003).
8. Morris, M. K., Saez-Rodriguez, J., Sorger, P. K. & Lauffenburger, D. A. Logic-based models for the analysis of cell signaling networks. *Biochemistry* **49,** 3216–3224 (2010).
9. Mayo, A. E., Setty, Y., Shavit, S., Zaslaver, A. & Alon, U. Plasticity of the cis-regulatory input function of a gene. *PLoS Biol.* **4,** e45 (2006).
10. Anderson, J. C., Voigt, C. A. & Arkin, A. P. Environmental signal integration by a modular AND gate. *Mol. Syst. Biol.* **3,** 133 (2007).
11. Guet, C. C., Elowitz, M. B., Hsing, W. & Leibler, S. Combinatorial synthesis of genetic networks. *Science* **296,** 1466–1470 (2002).
12. Rinaudo, K. *et al.* A universal RNAi-based logic evaluator that operates in mammalian cells. *Nature Biotechnol.* **25,** 795–801 (2007).
13. Weber, W. *et al.* A synthetic mammalian gene circuit reveals antituberculosis compounds. *Proc. Natl Acad. Sci. USA* **105,** 9994–9998 (2008).
14. Ellis, T., Wang, X. & Collins, J. J. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nature Biotechnol.* **27,** 465–471 (2009).
15. Lou, C. *et al.* Synthesizing a novel genetic sequential logic circuit: a push-on push-off switch. *Mol. Syst. Biol.* **6,** 350 (2010).
16. Tabor, J. J. *et al.* A synthetic genetic edge detection program. *Cell* **137,** 1272–1281 (2009).
17. Friedland, A. E. *et al.* Synthetic gene networks that count. *Science* **324,** 1199–1202 (2009).
18. Tan, C., Marguet, P. & You, L. Emergent bistability by a growth-modulating positive feedback circuit. *Nature Chem. Biol.* **5,** 842–848 (2009).
19. Scharle, T. W. Axiomatization of propositional calculus with Sheffer functors. *Notre Dame J. Formal Logic* **6,** 209–217 (1965).
20. Yokobayashi, Y., Weiss, R. & Arnold, F. H. Directed evolution of a genetic circuit. *Proc. Natl Acad. Sci. USA* **99,** 16587–16591 (2002).
21. Sneppen, K. *et al.* A mathematical model for transcriptional interference by RNA polymerase traffic in *Escherichia coli*. *J. Mol. Biol.* **346,** 399–409 (2005).
22. Pesci, E. C., Pearson, J. P., Seed, P. C. & Iglewski, B. H. Regulation of las and rhl quorum sensing in Pseudomonas aeruginosa. *J. Bacteriol.* **179,** 3127–3132 (1997).
23. Karig, D. & Weiss, R. Signal-amplifying genetic circuit enables *in vivo* observation of weak promoter activation in the Rhl quorum sensing system. *Biotechnol. Bioeng.* **89,** 709–718 (2005).
24. Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. Gene regulation at the single-cell level. *Science* **307,** 1962–1965 (2005).
25. Pedraza, J. M. & van Oudenaarden, A. Noise propagation in gene networks. *Science* **307,** 1965–1969 (2005).
26. Katz, R. H. & Borriello, G. *Contemporary Logic Design* 141–146 (Prentice Hall, 1994).
27. Danino, T., Mondragon-Palomino, O., Tsimring, L. & Hasty, J. A synchronized quorum of genetic clocks. *Nature* **463,** 326–330 (2010).
28. Clancy, K. & Voigt, C. A. Programming cells: towards an automated 'genetic compiler'. *Curr. Opin. Biotechnol.* **21,** 572–581 (2010).
29. Ilachinski, A. *Cellular Automata: A Discrete Universe* 1–18 (World Scientific, 2001).
30. Raju, B. S. & Mullick, S. K. Programmable cellular arrays. *Int. J. Control* **14,** 1041–1061 (1971).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank W. Mulyasasmita and K. Temme for critical discussions. This work was supported by the National Science Foundation (SynBERC, NSF#0943385 and NSF Sandpit CCF-0943385) and the Office of Naval Research.

Author Contributions A.T. designed and performed the experiments, analysed the data and wrote the manuscript. J.J.T. designed experiments and edited the manuscript. C.A.V. designed experiments, analysed the data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.A.V. (cavoigt@picasso.ucsf.edu).

LETTER

doi:10.1038/nature09570

Noise correlations improve response fidelity and stimulus encoding

Jon Cafaro[2] & Fred Rieke[1,2]

Computation in the nervous system often relies on the integration of signals from parallel circuits with different functional properties. Correlated noise in these inputs can, in principle, have diverse and dramatic effects on the reliability of the resulting computations[1–8]. Such theoretical predictions have rarely been tested experimentally because of a scarcity of preparations that permit measurement of both the covariation of a neuron's input signals and the effect on a cell's output of manipulating such covariation. Here we introduce a method to measure covariation of the excitatory and inhibitory inputs a cell receives. This method revealed strong correlated noise in the inputs to two types of retinal ganglion cell. Eliminating correlated noise without changing other input properties substantially decreased the accuracy with which a cell's spike outputs encoded light inputs. Thus, covariation of excitatory and inhibitory inputs can be a critical determinant of the reliability of neural coding and computation.

Differences in the properties of excitatory and inhibitory synaptic inputs to a target cell provide a key control of neural activity. Feedforward inhibitory synaptic input is a ubiquitous example. A delay in inhibitory input relative to excitatory input, for example by an extra synaptic delay in the circuit providing inhibitory input, can limit response duration to the time window in which the target cell receives excitatory but not inhibitory input[9]. More generally, inhibitory input can cancel unwanted responses by arriving before or at the same time as excitatory input[10–13]. Theoretical work illustrates how the effectiveness of these computations depends on the strength of covariation between excitatory and inhibitory synaptic inputs[8]. Thus, although synaptic noise will always decrease the reliability of the neural response, strong noise correlations, unlike independent noise, could allow fluctuations in inhibitory synaptic input to cancel corresponding fluctuations in excitatory synaptic input[5] (Fig. 1). Such noise correlations can arise if noise within excitatory and inhibitory pathways originates from a common source (Fig. 1, left), for example in densely and randomly connected recurrent networks[14]. Noise cancellation in synaptic integration could in turn reduce trial-to-trial variability in a cell's spike output (Fig. 1, right).

The extent and impact of noise correlations depends on several network and cellular properties, including nonlinearities in synaptic transmission[15] or spike generation[16] that could decrease correlation strength. This dependence makes it difficult to predict the importance of noise correlations from modelling alone or from correlations measured in cell pairs. Work on the retina provides a rare opportunity to provide quantitative experimental information about how noise correlations affect the coding of physiologically relevant stimuli. Our goal was first to measure covariation of the excitatory and inhibitory synaptic inputs received by a retinal ganglion cell (Fig. 1, (Q1)) and then to test how these noise correlations affect the encoding of light stimuli in a cell's spike output (Fig. 1, (Q2)).

Quantifying the covariation of excitatory and inhibitory synaptic input requires measuring these two conductances simultaneously or near simultaneously. To do this, we rapidly alternated the ganglion cell voltage between the reversal potentials for excitatory and inhibitory synaptic inputs, collecting a single sample of each input every 10 ms (Fig. 2a). Control experiments indicated that the voltage at the synaptic receptors had reached a near-constant value at these sampling times (Supplementary Fig. 1). This sampling rate is high in comparison with the 50–100 ms time course of a ganglion cell's response to light inputs. To check how well this procedure captured light-dependent changes in conductance, we compared the simultaneously measured conductances with those measured non-simultaneously when the voltage was held constant at the excitatory or inhibitory reversal potential. Mean excitatory and inhibitory conductances resulting from a repeated, modulated light input differed minimally (Fig. 2b). In 21 cells, the alternating voltage approach captured 99.9 ± 0.6% of the power of the conductance signal and 83 ± 4% of that of the conductance noise (mean ± s.e.m.; see Methods). Thus, simultaneous conductance measurements capture most of the structure in the synaptic inputs a ganglion cell receives.

Simultaneous conductances measured during constant light input often exhibited spontaneous excitatory synaptic events accompanied in time by inhibitory synaptic events (Fig. 2c, top, black arrowheads). Such events in fact typically occurred together. Correlated noise events were rarely observed during non-simultaneously measured conductances (Fig. 2c, bottom). Correspondingly, the cross-correlation function for

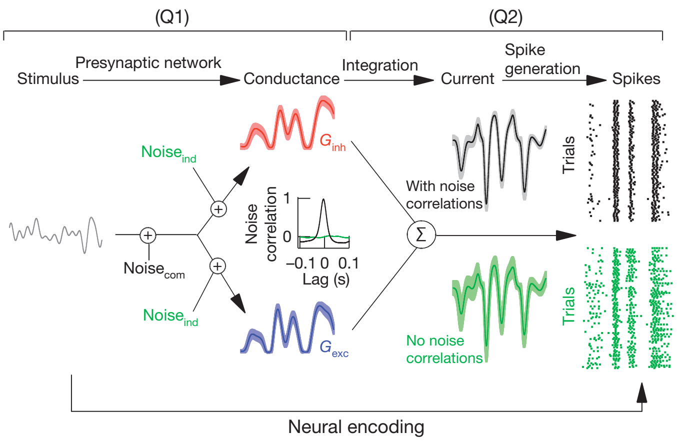


Figure 1 | Effects of noise correlations on the variability of synaptic current and spike output. Neural encoding consists of three basic steps: a stimulus shapes excitatory (blue, G_{exc}) and inhibitory (red, G_{inh}) synaptic conductances; these conductances then shape synaptic currents; and the resulting currents control spike generation to produce a sequence of action potentials (spikes). Noise correlations will be strong if a common source dominates noise in excitatory and inhibitory pathways (Noise$_{com}$) and minimal if the dominant noise source arises independently (Noise$_{ind}$). Correlated (black traces) as opposed to uncorrelated (green traces) noise between excitatory and inhibitory conductances can lead to lower variability of both the synaptic current and the spike output (shaded regions around traces). Understanding this issue requires answering two questions. (Q1) How much do converging excitatory and inhibitory input covary? (Q2) What is the impact of such noise correlations on the neural output?

[1]Howard Hughes Medical Institute. University of Washington, Seattle, Washington 98195, USA. [2]Department of Physiology and Biophysics, University of Washington, Seattle, Washington 98195, USA.

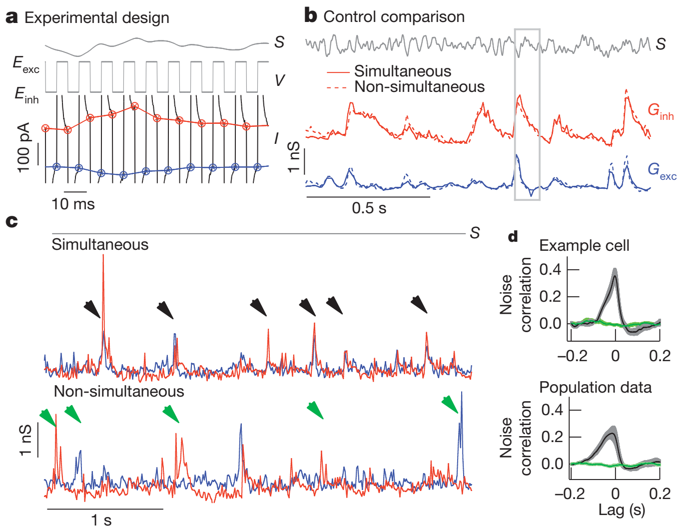


Figure 2 | Near-simultaneous recording of excitatory and inhibitory synaptic input to an ON–OFF directionally selective ganglion cell. a, Light stimulus (*S*) is presented while the voltage (*V*) of the cell alternates between the excitatory (E_{exc}) and inhibitory (E_{inh}) reversal potentials. Excitatory (blue) and inhibitory (red) synaptic currents (*I*) are sampled at the end of each voltage step. **b**, Conductances derived from measured currents (Methods) and averaged across multiple repeats of the same stimulus (*S*). Simultaneously measured conductances (solid lines) closely match those (dashed lines) measured non-simultaneously with the voltage held fixed at the reversal potentials for excitatory or inhibitory input (both excitatory and inhibitory correlations are 0.91 ± 0.01 (mean ± s.e.m.), 21 cells). **a** is an enlarged view of the boxed region of **b**. **c**, Top: section of simultaneously recorded conductances during constant light input shows correlated excitatory and inhibitory spontaneous events (black arrowheads). Bottom: non-simultaneously recorded conductances also show spontaneous events (green arrowheads), but they are rarely correlated. Records have been resampled at 50 Hz for comparison with the top conductances. **d**, Top: cross-correlation (mean ± s.e.m., 10 trials) of excitatory and inhibitory conductances in an example cell during simultaneous (black) and non-simultaneous (green) recording. Bottom, cross-correlation for all recorded cells (mean ± s.e.m., 6 cells).

simultaneously measured excitatory and inhibitory conductances during constant light input showed considerable structure, unlike the cross-correlation for non-simultaneously measured conductances (single cell: Fig. 2d, top; population: Fig. 2d, bottom). Thus, simultaneous conductance recordings revealed correlations between converging synaptic inputs that were inaccessible from more conventional recordings.

To determine both the strength of noise correlations during modulated light input and their effect on a cell's spike output, we first considered midget ganglion cells, which comprise the majority of ganglion cells in the primate retina[17]. Midget ganglion cells receive delayed feed-forward synaptic inhibition, where the delay reflects an extra synapse in the circuit controlling inhibitory input. Thus, excitatory input comes directly from bipolar cells, whereas inhibitory input comes from amacrine cells that themselves receive input from bipolar cells[18]. Similar delayed feed-forward inhibition is a characteristic of many cortical circuits, including hippocampus, cerebellum, barrel cortex and auditory cortex[9,11,12,19–21]. We simultaneously recorded excitatory and inhibitory synaptic inputs during a full-field modulated light stimulus (Fig. 3a, left) and estimated variability in the synaptic responses by subtracting the average synaptic input from each individual trial (Fig. 3a, right). The peak correlation strength of the resulting residuals ranged from 0.15 to 0.5 (Fig. 3b, black traces). Noise correlations in the interleaved non-simultaneous conductances were substantially smaller (Fig. 3b, green traces). Slow drift in the light response accounted for the remaining noise correlations in the non-simultaneous conductances (Supplementary Fig. 2).

The alternating-voltage technique could produce artefactual noise correlations by overshooting the appropriate reversal potentials for excitatory or inhibitory synaptic inputs. For example, holding at a voltage positive relative to the excitatory reversal potential could cause

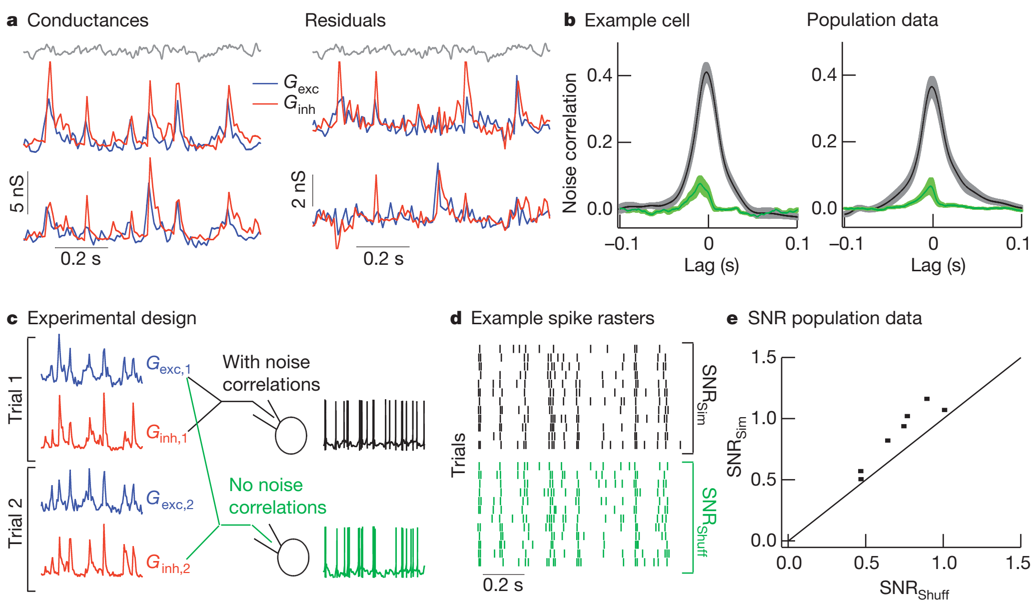


Figure 3 | Strength and impact of noise correlations in synaptic inputs to primate midget ganglion cells. a, Left: two trials of simultaneously recorded conductances during modulated light input (grey). Right: residual conductances (trials from left with mean subtracted), which estimate noise in each trial. **b**, Left: cross-correlation (mean ± s.e.m., 12 trials) of excitatory and inhibitory residual conductances in an example cell during simultaneous (black) and non-simultaneous (green) recording. Right: cross-correlation for all recorded cells (mean ± s.e.m., 15 cells). **c**, Logic of dynamic-clamp experiments using simultaneously or shuffled simultaneous conductances in place of synaptic input. **d**, Example spike trains from 12 dynamic-clamp trials of simultaneous conductances (black) or their shuffled counterparts (green). SNR, signal-to-noise ratio. **e**, Signal-to-noise ratio of spike trains generated from simultaneous conductances versus that of spike trains generated from shuffled conductances (dots). The signal-to-noise ratio for simultaneous conductances was 1.22 ± 0.04 times higher than that for shuffled conductances (mean ± s.e.m., 7 cells, $P = 0.0015$).

an increase in the excitatory conductance to be misinterpreted as an increase in both the excitatory and inhibitory conductances, thus leading to an artefactual correlation. A similar logic holds if a cell is held more negative than the reversal potential for inhibitory input. However, if anything the alternating-voltage technique fell short of the actual reversal potentials and hence underestimated the strength of noise correlations (Supplementary Fig. 3).

To determine the effect of covariation of excitatory and inhibitory synaptic inputs on a midget ganglion cell's response to physiological inputs, we compared the pattern of spikes produced by simultaneous (with noise correlations) and non-simultaneous (without noise correlations) conductances in dynamic-clamp experiments (Fig. 3c and Supplementary Fig. 4). The non-simultaneous conductances consisted of shuffled pairings of simultaneously recorded excitatory and inhibitory conductances; this procedure removed noise correlations while holding all other statistics constant. We compared the precision of the spike responses to the two sets of conductances by calculating the signal-to-noise ratio from repeated dynamic-clamp trials (Fig. 3d; see Methods). In all cases, the signal-to-noise ratio was higher for conductances with noise correlations (Fig. 3e). Quantifying the temporal precision of the spike responses using a spike distance metric[22,23] gave similar results (data not shown). Thus, the precision of a midget cell's output in response to light stimuli depends on the covariation of excitatory and inhibitory synaptic inputs.

Feed-forward synaptic inhibition can serve a more diverse functional role when the amplitude or timing of inhibitory input relative to excitatory input depends on the stimulus. For example, the ability of a subset of retinal ganglion cells to respond to the direction of a moving object[24,25] (Fig. 4a, b) relies on cancellation of excitatory input by inhibitory input in the non-preferred direction[10]. Covariation of excitatory and inhibitory synaptic inputs could make such a mechanism robust to noise, for example by preventing a larger-than-average excitatory synaptic event from overwhelming the corresponding inhibitory synaptic event and causing a response to movement in an inappropriate direction. To test this proposal, we recorded simultaneous conductances in mouse ON–OFF directionally selective ganglion cells (ON–OFF DSGCs) in response to a bar of light moving in different directions (Fig. 4a). Excitatory and inhibitory conductances showed strong noise correlations that were largely absent from non-simultaneous conductances (Fig. 4d; see Supplementary Fig. 5 for results from full-field light stimuli). Both excitatory and inhibitory conductances and the strength of the noise correlations depended on bar direction (Fig. 4c, d). For example, noise correlations in the non-preferred direction were three to four times stronger than those in the preferred direction. Furthermore, excitatory and inhibitory conductances showed near-perfect covariation in the non-preferred direction.

We tested the impact of noise correlations on direction tuning using simultaneous (with noise correlations) and non-simultaneous (without noise correlations) conductances in dynamic-clamp experiments; non-simultaneous conductances consisted of simultaneous conductances shuffled between trials but not bar directions. Both the mean and the standard deviation of the firing rate in the non-preferred direction were considerably higher for non-simultaneous conductances (Fig. 4e, f). The failure of a cell to attenuate its response reliably for movement in the non-preferred direction should negatively affect its ability to encode direction. Indeed, each recorded cell showed greater direction selectivity for the simultaneous conductances (Fig. 4g). Thus, the computation underlying directional selectivity depends on covariation of excitatory and inhibitory synaptic inputs and the resulting cancellation of noise shared between the circuits providing each type of input.

Computation in the retina follows a basic plan found in many other neural circuits: signals in a common population of inputs diverge to parallel and functionally dissimilar pathways, and integration of the signals from multiple parallel pathways governs the output of the circuit. Divergence into separate excitatory and inhibitory circuits is a prominent example of such a motif. Noise in shared inputs naturally causes covariation of signals in the parallel pathways. The strength of such noise correlations will depend on cellular properties within the network[15,16],

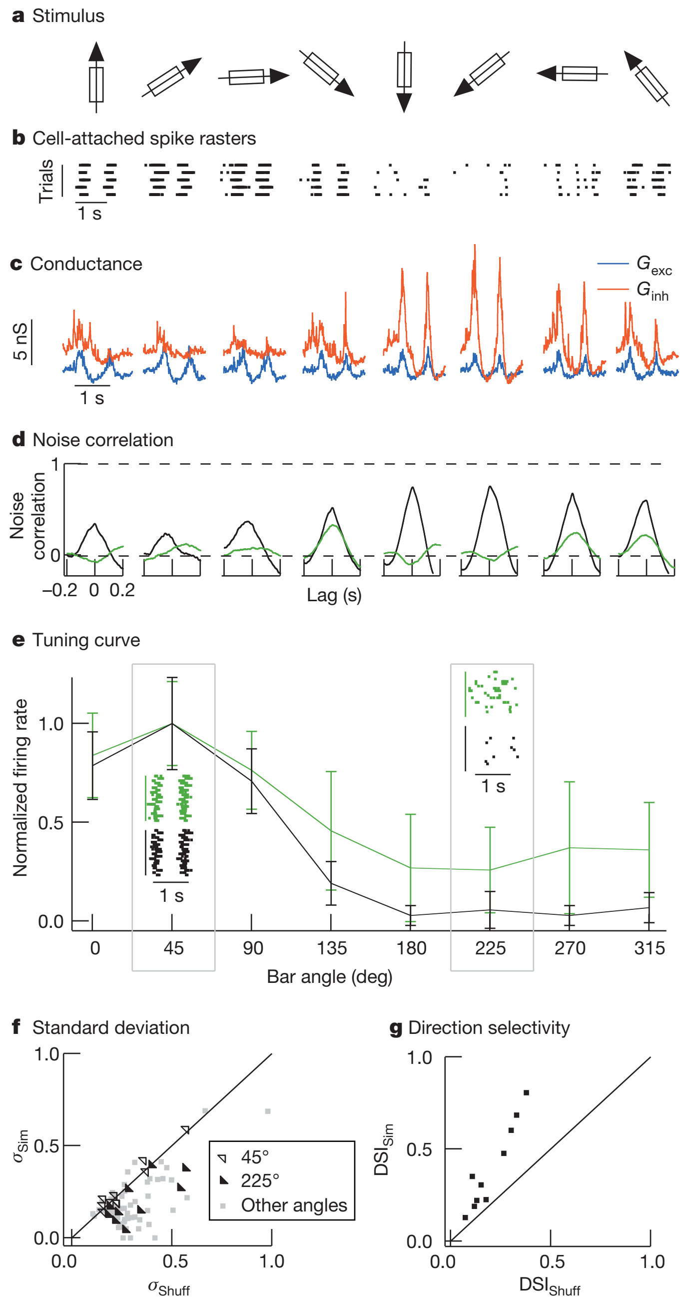


Figure 4 | Strength and impact of noise correlations in synaptic inputs to ON–OFF directionally selective ganglion cells. a, A bar of light was moved in eight directions, at 45° increments in random order. **b**, Extracellular (cell-attached configuration) spike responses to the moving bar. **c**, Examples of simultaneously recorded conductances showing tuning of excitatory (blue) and inhibitory (red) conductances. **d**, Simultaneous conductances (black) show strong noise correlations that are largely absent from the non-simultaneous conductances (green). **e**, Normalized directional tuning (spike count versus direction) from a single dynamic-clamp experiment (mean ± s.d.) for 20 trials of simultaneous or shuffled simultaneous conductances. Insets at 45° (preferred direction) and 225° (non-preferred direction) show spike rasters. **f**, Standard deviation of the normalized spike count is significantly smaller for simultaneous trials than for shuffled trials in non-preferred directions (135–315°; $P < 0.05$, 10 cells). Standard deviations in the preferred direction were similar. **g**, Direction selectivity index (DSI; see Methods) is 2.0 ± 0.2 times larger for simultaneous conductances than for shuffled conductances (mean ± s.e.m., 10 cells, $P = 0.0002$).

the stimulus delivered[26] (Fig. 4) and the state of the network[27]. Thus, excitatory and inhibitory inputs to cells in some, but not all, circuits are expected to show strong noise correlations, as indeed is the case in barrel cortex[27,28]. Here we put such noise correlations in the context of the coding of physiologically relevant stimuli. Our results reveal a critical role for noise correlations in maintaining appropriate cancellation of excitatory and inhibitory inputs and thus sharpening tuning to specific stimuli. This work provides an example of neurons that perform computations reliant on noise correlations. Given the prevalence of circuits in which feed-forward inhibition shapes neural responses[9,11,12,19–21], noise correlations probably have a similar role in other neural circuits.

METHODS SUMMARY

We took electrical recordings from midget ganglion cells in primate and ON–OFF DSGCs in mouse retinas using patch-clamp techniques as previously described[23,29]. Light stimuli were delivered from light-emitting diodes or an organic light-emitting diode monitor (eMagin). Mean light levels for all experiments were near 5,000 absorbed photons per cone per second.

The 10-ms cycle period during the simultaneous conductance recordings allows us to resolve input at 50 Hz and below. The fraction of the measured current variance at this cycle time was determined by calculating the fraction of the variance of the non-simultaneous (constant-voltage) conductances that can be accounted for by the variance of the simultaneous conductances.

Signal-to-noise ratios of spike outputs were calculated by forming spike trains of zeroes and ones from each trial, with 1-ms resolution. The mean and trial residuals of these spike trains were calculated and the power spectra of these functions were assessed and corrected for sample number bias[30]. Power spectra were integrated between 1 and 20 Hz and the result for the mean responses was divided by that for the residuals (Supplementary Fig. 6).

Spike number in ON–OFF DSGCs in response to the moving bar was summed over the entire duration of the bar's movement. The direction selectivity index[10] was calculated as DSI $= |\sum \mathbf{v}_i / \sum \boldsymbol{r}_i|$, where $\mathbf{v}_i$ are vectors of lengths $\boldsymbol{r}_i$, equal to the normalized firing rate, and point in the direction of the moving bar that produced the presented conductances.

Current injected into a cell (I) during dynamic-clamp experiments[31] was calculated as

$$I(t) = G_{\mathrm{exc}}(t)(V(t-\Delta t) - E_{\mathrm{exc}}) + G_{\mathrm{inh}}(t)(V(t-\Delta t) - E_{\mathrm{inh}})$$

where G_{exc} and G_{inh} are a pair of conductances recorded during light stimulation, V is the cell's membrane potential, and E_{exc} and E_{inh} are reversal potentials set respectively at 0 mV and −80 mV. Changing the inhibitory reversal potential, E_{inh}, to −50 mV did not substantially affect the results.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 July; accepted 11 October 2010.
Published online 5 December 2010.

1. Shadlen, M. N. & Newsome, W. T. Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.* **4,** 569–579 (1994).
2. Softky, W. R. & Koch, C. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci.* **13,** 334–350 (1993).
3. Abbott, L. F. & Dayan, P. The effect of correlated variability on the accuracy of a population code. *Neural Comput.* **11,** 91–101 (1999).
4. Romo, R., Hernandez, A., Zainos, A. & Salinas, E. Correlated neuronal discharges that increase coding efficiency during perceptual discrimination. *Neuron* **38,** 649–657 (2003).
5. Salinas, E. & Sejnowski, T. J. Impact of correlated synaptic input on output firing rate and variability in simple neuronal models. *J. Neurosci.* **20,** 6193–6209 (2000).
6. Dan, Y., Alonso, J. M., Usrey, W. M. & Reid, R. C. Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nature Neurosci.* **1,** 501–507 (1998).
7. Nirenberg, S., Carcieri, S. M., Jacobs, A. L. & Latham, P. E. Retinal ganglion cells act largely as independent encoders. *Nature* **411,** 698–701 (2001).
8. Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population coding and computation. *Nature Rev. Neurosci.* **7,** 358–366 (2006).
9. Pouille, F. & Scanziani, M. Enforcement of temporal fidelity in pyramidal cells by somatic feed-forward inhibition. *Science* **293,** 1159–1163 (2001).
10. Taylor, W. R. & Vaney, D. I. Diverse synaptic mechanisms generate direction selectivity in the rabbit retina. *J. Neurosci.* **22,** 7712–7720 (2002).
11. Wilent, W. B. & Contreras, D. Dynamics of excitation and inhibition underlying stimulus selectivity in rat somatosensory cortex. *Nature Neurosci.* **8,** 1364–1370 (2005).
12. Wehr, M. & Zador, A. M. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* **426,** 442–446 (2003).
13. Leary, C. J., Edwards, C. J. & Rose, G. J. Midbrain auditory neurons integrate excitation and inhibition to generate duration selectivity: an *in vivo* whole-cell patch study in anurans. *J. Neurosci.* **28,** 5481–5493 (2008).
14. Renart, A. *et al.* The asynchronous state in cortical circuits. *Science* **327,** 587–590 (2010).
15. Trong, P. K. & Rieke, F. Origin of correlated activity between parasol retinal ganglion cells. *Nature Neurosci.* **11,** 1343–1351 (2008).
16. de la Rocha, J., Doiron, B., Shea-Brown, E., Josic, K. & Reyes, A. Correlation between neural spike trains increases with firing rate. *Nature* **448,** 802–806 (2007).
17. Dacey, D. M. & Petersen, M. R. Dendritic field size and morphology of midget and parasol ganglion cells of the human retina. *Proc. Natl Acad. Sci. USA* **89,** 9666–9670 (1992).
18. Calkins, D. J. & Sterling, P. Absence of spectrally specific lateral inputs to midget ganglion cells in primate retina. *Nature* **381,** 613–615 (1996).
19. Gabernet, L., Jadhav, S. P., Feldman, D. E., Carandini, M. & Scanziani, M. Somatosensory integration controlled by dynamic thalamocortical feed-forward inhibition. *Neuron* **48,** 315–327 (2005).
20. Luna, V. M. & Schoppa, N. E. GABAergic circuits control input-spike coupling in the piriform cortex. *J. Neurosci.* **28,** 8851–8859 (2008).
21. Mittmann, W., Koch, U. & Hausser, M. Feed-forward inhibition shapes the spike output of cerebellar Purkinje cells. *J. Physiol. (Lond.)* **563,** 369–378 (2005).
22. Victor, J. D. & Purpura, K. P. Nature and precision of temporal coding in visual cortex: a metric-space analysis. *J. Neurophysiol.* **76,** 1310–1326 (1996).
23. Murphy, G. J. & Rieke, F. Network variability limits stimulus-evoked spike timing precision in retinal ganglion cells. *Neuron* **52,** 511–524 (2006).
24. Barlow, H. B., Hill, R. M. & Levivk, W. R. Retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit. *J. Physiol. (Lond.)* **173,** 377–407 (1964).
25. Weng, S., Sun, W. & He, S. Identification of ON-OFF direction-selective ganglion cells in the mouse retina. *J. Physiol. (Lond.)* **562,** 915–923 (2005).
26. Cohen, M. R. & Newsome, W. T. Context-dependent changes in functional circuitry in visual area MT. *Neuron* **60,** 162–173 (2008).
27. Gentet, L. J., Avermann, M., Matyas, F., Staiger, J. F. & Petersen, C. C. Membrane potential dynamics of GABAergic neurons in the barrel cortex of behaving mice. *Neuron* **65,** 422–435 (2010).
28. Okun, M. & Lampl, I. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature Neurosci.* **11,** 535–537 (2008).
29. Dunn, F. A., Lankheet, M. J. & Rieke, F. Light adaptation in cone vision involves switching between receptor and post-receptor sites. *Nature* **449,** 603–606 (2007).
30. van Hateren, J. H. & Snippe, H. P. Information theoretical evaluation of parametric models of gain control in blowfly photoreceptor cells. *Vision Res.* **41,** 1851–1865 (2001).
31. Sharp, A. A., O'Neil, M. B., Abbott, L. F. & Marder, E. Dynamic clamp: computer-generated conductances in real neurons. *J. Neurophysiol.* **69,** 992–995 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Dacey, O. Packer, J. Crook, B. Peterson and T. Haun for providing primate tissue; P. Newman and E. Martinson for technical assistance; T. Azevedo, E. J. Chichilnisky, F. Dunn, G. Murphy, S. Kuo, E. Shea-Brown, M. Shadlen and W. Spain for comments on the manuscript and discussions. Support was provided by HHMI and NIH (EY-11850).

Author Contributions J.C. and F.R. designed and carried out the experiments, J.C. analysed the data and J.C. and F.R. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.R. (rieke@u.washington.edu).

METHODS

Electrical recordings were made from midget ganglion cells in primate and ON–OFF DSGCs in mouse retinas as previously described[23,29]. Midget ganglion cells were identified by their relatively sustained response to light steps and characteristic morphology[17,29,32]. ON–OFF DSGCs were identified by a combination of at least two of the following criteria: an on–off light response to a brief light step, a bistratified morphology and a directionally selective spike response.

Light stimuli were delivered from light-emitting diodes or an organic light-emitting diode monitor (eMagin). Mean light levels for all experiments were near 5000 absorbed photons per cone per second. Full-field stimuli consisted of 10 s of constant light followed by 10 s of 50%-contrast modulated light (low-pass-filtered at 60 Hz) repeated for 5–20 trials. Moving bars were 180 µm wide, 720 µm long, moved at 864 µm s^{-1} along the long axis and had a contrast of between 100 and 150%.

For all recordings a flat-mounted piece of retina was superfused with warmed (31–34 °C) and oxygenated (5% CO_2, 95% O_2) Ames solution. Midget cell dynamic-clamp experiments were performed with receptors mediating excitatory and inhibitory synaptic input blocked (10 µM NBQX, 1 µM stychnine, 10 µM gabazine). Pipettes for voltage-clamp recordings were filled with a Cs-based internal solution (105 mM $CsCH_3SO_3$, 10 mM TEA-Cl, 20 mM HEPES, 10 mM EGTA, 5 mM Mg-ATP, 0.5 mM Tris-GTP and 2 mM QX-314, pH ~7.3, ~280 mosM). Pipettes for dynamic-clamp experiments were filled with a K-based internal solution (110 mM K aspartate, 1 mM MgCl, 10 mM HEPES, 5 mM NMDG, 0.5 mM $CaCl_2$, 10 mM phosphocreatine, 4 mM Mg-ATP and 0.5 mM Tris-GTP, pH ~7.2, ~280 mosM). Liquid junction potentials were ~10 mV and were not compensated throughout the text. Low access resistance was critical, and only cells with access resistance below 20 MΩ were included for analysis. Access resistance was partially compensated for (75% for experiments using an Axopatch 200B amplifier; 50% compensation and prediction for experiments using a Multiclamp 700B amplifier). Conductances were derived from excitatory and inhibitory synaptic currents by dividing the currents by assumed driving forces corresponding to voltages of −62 and +62 mV, respectively.

Both ganglion cell types showed evidence for NMDA-receptor-mediated conductances (J-shaped I–V plots that became linear in the presence of 10 µM APV). The presence of an NMDA conductance could cause noise correlations to be substantially underestimated if the voltage is substantially below the excitatory reversal potential. However, we observed only a weak impact of this conductance when noise correlations were compared before and after application of APV. Results from two cells recorded only in the presence of APV were included in the full data set.

The 10-ms cycle period during the simultaneous conductance recordings allows us to resolve input at 50 Hz and below. The fraction of the measured current variance at this cycle time was determined by calculating the fraction of the variance of the non-simultaneous (constant-voltage) conductances that can be accounted for by the variance of the simultaneous conductances.

Signal-to-noise ratios of spike outputs were calculated by forming spike trains of zeroes and ones from each trial, with 1-ms resolution. The mean and trial residuals of these spike trains were calculated and the power spectra of these functions were assessed and corrected for sample number bias[30]. Power spectra were integrated between 1 and 20Hz and the result for the mean responses was divided by that for the residuals (Supplementary Fig. 6).

Spike number in ON–OFF DSGCs in response to the moving bar was summed over the entire duration of the bar's movement. The direction selectivity index[10] was calculated as $\mathrm{DSI} = |\sum \mathbf{v}_i / \sum \boldsymbol{r}_i|$, where $\mathbf{v}_i$ are vectors of lengths $\boldsymbol{r}_i$, equal to the normalized firing rate, and point in the direction of the moving bar that produced the presented conductances.

Current injected into a cell (I) during dynamic-clamp experiments[31] was calculated as

$$I(t) = G_{\mathrm{exc}}(t)(V(t-\Delta t) - E_{\mathrm{exc}}) + G_{\mathrm{inh}}(t)(V(t-\Delta t) - E_{\mathrm{inh}})$$

where G_{exc} and G_{inh} are a pair of conductances recorded during light stimulation, V is the cell's membrane potential, and E_{exc} and E_{inh} are reversal potentials set respectively at 0 mV and −80 mV. Changing the inhibitory reversal potential, E_{inh}, to −50 mV did not substantially affect the results.

Correlations were calculated using the 'xcov' function in MATLAB, release 2009a (MathWorks) and normalized using the 'coef' option. Briefly, this function calculates the cross-correlation after subtracting the means from each trial and normalizes by the geometric mean of the autocorrelation (see Supplementary Information, equation (2.1)).

32. Polyak, S. & Willmer, E. N. Retinal structure and colour vision. *Doc. Ophthalmol.* **3,** 24–56 (1949).

LETTER

doi:10.1038/nature09573

Crystal structure of bacterial RNA polymerase bound with a transcription inhibitor protein

Shunsuke Tagami[1,2], Shun-ichi Sekine[1,2,3], Thirumananseri Kumarevel[4]†, Nobumasa Hino[2], Yuko Murayama[1,2], Syunsuke Kamegamori[1,2], Masaki Yamamoto[5], Kensaku Sakamoto[2] & Shigeyuki Yokoyama[1,2,3]

The multi-subunit DNA-dependent RNA polymerase (RNAP) is the principal enzyme of transcription for gene expression. Transcription is regulated by various transcription factors. Gre factor homologue 1 (Gfh1), found in the *Thermus* genus, is a close homologue of the well-conserved bacterial transcription factor GreA, and inhibits transcription initiation and elongation by binding directly to RNAP[1–8]. The structural basis of transcription inhibition by Gfh1 has remained elusive, although the crystal structures of RNAP and Gfh1 have been determined separately[6–9]. Here we report the crystal structure of *Thermus thermophilus* RNAP complexed with Gfh1. The amino-terminal coiled-coil domain of Gfh1 fully occludes the channel formed between the two central modules of RNAP; this channel would normally be used for nucleotide triphosphate (NTP) entry into the catalytic site. Furthermore, the tip of the coiled-coil domain occupies the NTP β-γ phosphate-binding site. The NTP-entry channel is expanded, because the central modules are 'ratcheted' relative to each other by ~7°, as compared with the previously reported elongation complexes. This 'ratcheted state' is an alternative structural state, defined by a newly acquired contact between the central modules. Therefore, the shape of Gfh1 is appropriate to maintain RNAP in the ratcheted state. Simultaneously, the ratcheting expands the nucleic-acid-binding channel, and kinks the bridge helix, which connects the central modules. Taken together, the present results reveal that Gfh1 inhibits transcription by preventing NTP binding and freezing RNAP in the alternative structural state. The ratcheted state might also be associated with other aspects of transcription, such as RNAP translocation and transcription termination.

RNAP synthesizes RNA complementary to the template DNA (Supplementary Fig. 1a). Crystallographic studies of RNAPs from thermophilic bacteria and RNAP II (Pol II) from the yeast *Saccharomyces cerevisiae* have revealed the overall structure of RNAP, which resembles a crab's claw[9–11] (Supplementary Fig. 1b). In the transcribing RNAP (elongation complex, EC), the nascent RNA strand remains bound to the template DNA strand, forming an 8−9 base pair DNA•RNA hybrid. The DNA•RNA hybrid and the downstream DNA duplex are tightly held in the 'primary channel' formed between the pincers of the crab claw, in the EC structures of the *S. cerevisiae* and *T. thermophilus* RNAPs[12–21]. The catalytic site of nucleotide addition resides at the joint of the pincers. The substrate NTP is considered to enter the catalytic site through a pore (the 'secondary channel') on the back side of the crab claw (Supplementary Fig. 1b). The two pincers are connected by a long α-helix (the bridge helix), which is located at the junction of the DNA•RNA hybrid-binding site, the downstream DNA-binding site, and the secondary channel. The bridge helix is inherently flexible, adopting both the continuously-helical and kinked conformations[9–11,22,23].

In a previous study[24], we successfully crystallized *T. thermophilus* RNAP together with DNA, RNA and Gfh1, and collected X-ray diffraction data sets for two $P2_1$ crystals (crystals 1 and 2). The nucleic-acid scaffolds employed for the crystallization included the downstream duplex DNA, the DNA•RNA hybrid, and an upstream RNA hairpin 10 or 11 nucleotides (nt) from the RNA 3′ end (Supplementary Text 1, Supplementary Fig. 2a)[24]. In the present study, we determined the structures of the quaternary complex of RNAP•DNA•RNA•Gfh1 (EC•Gfh1) (Fig. 1a, b, Supplementary Table 1, and Methods). The structures of the three independent RNAP molecules in the asymmetric units of crystals 1 and 2 are all similar to each other (Supplementary Text 2, Supplementary Figs 3−5). RNAP and Gfh1 showed clear electron densities (Supplementary Fig. 6), and thus the inhibition mechanisms of Gfh1 were unambiguously revealed, as described below. In contrast, the electron densities of both the DNA and RNA were weak, so we only built the partial models (Supplementary Text 3, Supplementary Figs 2, 7).

The *S. cerevisiae* Pol II structure consists of four rigid modules, 'core', 'shelf', 'clamp' and 'jaw-lobe', which are mobile relative to each other[10]. The rigid modules of the bacterial RNAP were defined previously, on the basis of the structures of the *Thermus aquaticus* core enzyme and the *T. thermophilus* holoenzyme[11,25,26] (Supplementary Text 4). However, the conformations of the *T. thermophilus* RNAP in the present EC•Gfh1 structures differ appreciably from those in the previously reported structures of the core enzyme, the holoenzyme and ECs[13,14] (Fig. 1c, d). The large conformational differences enabled us to redefine the rigid-body modules of *T. thermophilus* RNAP (Fig. 1 a, b, Supplementary Table 2, Supplementary Text 4, Supplementary Figs 8, 9). These rigid modules include the 'core', 'shelf', and 'clamp' modules, which generally correspond well to those in Pol II. The exceptions are that the active-site and dock domains belong to the 'shelf' module, rather than to the 'core' module, and the β-domain 1 (or the protrusion domain in Pol II) and the β flap domain (or the wall domain in Pol II) are not included in the 'core' module in *T. thermophilus* RNAP (Supplementary Text 4, Supplementary Fig. 8). The core and shelf modules are the main structural elements that form the primary, secondary and RNA-exit channels (Fig. 1a, b, Supplementary Text 4). Gfh1 is accommodated in the secondary channel and its exterior region (Fig. 1a, b, Supplementary Fig. 6). The N-terminal domain (NTD) of Gfh1 forms a coiled coil and is inserted into the secondary channel. The major relative movements among the rigid-body modules are the 'ratcheting' between the shelf and core modules and the 'swinging' of the clamp relative to the shelf module (Fig. 1c, d, Supplementary Movies 1, 2), as described below in more detail.

The clamp module is connected to the shelf module by four loops (switches 1, 2, 4 and 5, Supplementary Table 2). The protruded clamp module swings relative to the shelf module by about 15° around the centre near switch 5, and is further tilted by about 5° in EC•Gfh1 (Fig. 1a, b, Supplementary Fig. 10, Supplementary Movie 2). Consequently, the region of the primary channel outside the hybrid-binding site is

[1]Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [2]RIKEN Systems and Structural Biology Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan. [3]Laboratory of Structural Biology, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [4]Structural and Molecular Biology Laboratory, RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan. [5]SR Life Science Instrumentation Unit, RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan. †Present address: Biometal Science Laboratory, RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan.

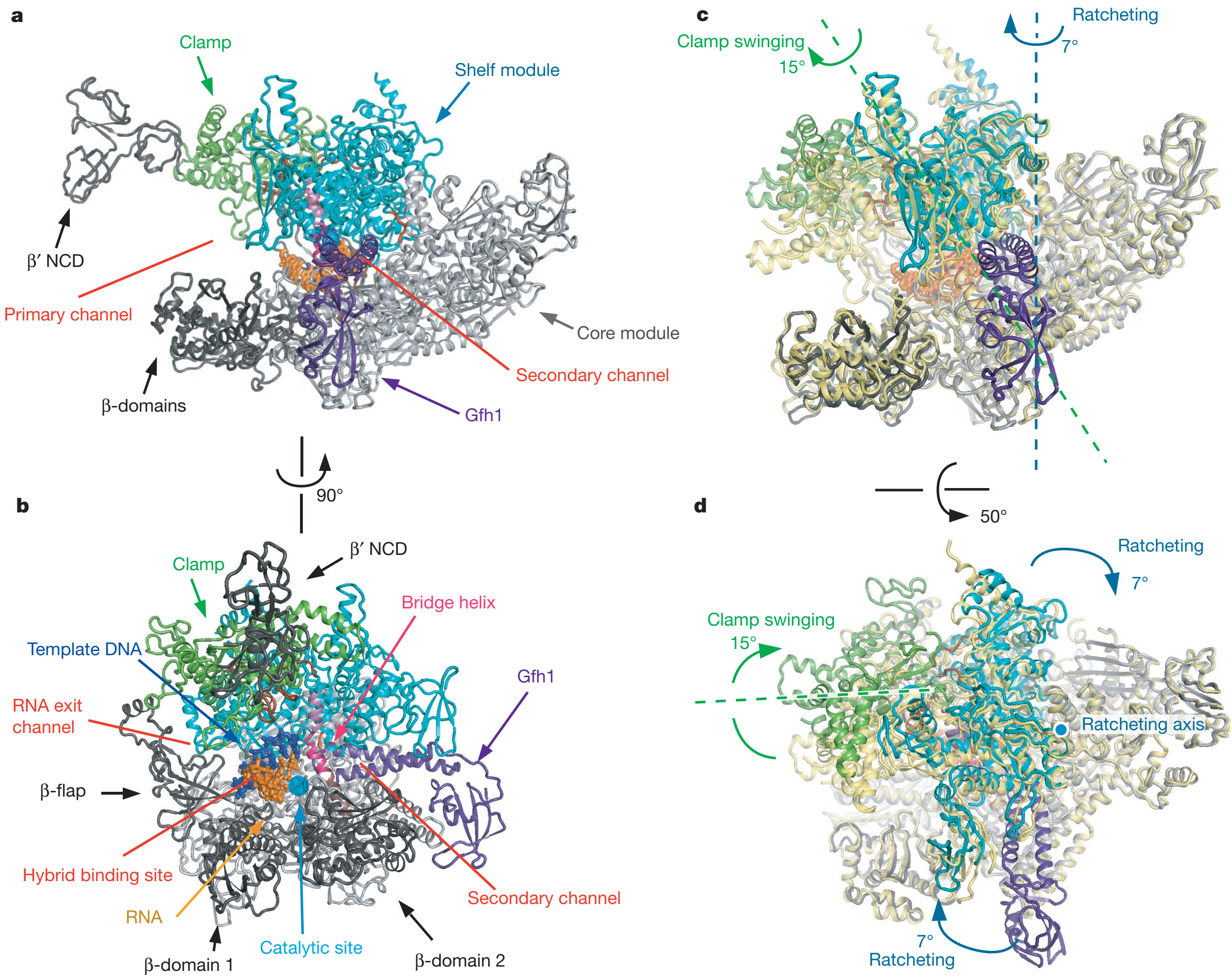


Figure 1 | Structure of EC•Gfh1. **a**, **b**, Overall structure of *T. thermophilus* EC•Gfh1 in two orientations. **c**, **d**, Superposition of EC•Gfh1 and EC (PDB 2O5I, yellow). The core modules of the two structures are superposed minimizing the root mean square deviation (RMSD) between Cα atoms. Two orientations are shown. The same colour scheme is used in all figures (RNA, orange; template DNA, blue; Gfh1, purple; shelf module, cyan; core module, grey; clamp, green; switches 1−5, brown; hinge loop, red; other domains, dark grey). The three regions of the bridge helix are coloured differently (N-terminal, dark pink; central, hot pink; C-terminal, violet).

widened (Supplementary Text 5, Supplementary Fig. 11). Considering that Gfh1 binding to RNAP occurs on the opposite side of the clamp module, and that Gfh1 does not directly contact the clamp module, the clamp swinging seems to be related to the hairpin structure in the RNA of the nucleic acid scaffolds (Supplementary Text 5).

The central body of RNAP is composed of the core and shelf modules, which are rotated by $\sim 7^{\circ}$ relative to each other in the present RNAP structures, as compared with the previously-reported EC state[13,14] (Figs 1c, d, 2a, and Supplementary Movie 1). We designate this novel RNAP state as the 'ratcheted' state, and the previous EC state as the 'tight' state. The shelf module is attached to the backboard part of the core module (Fig. 1a), through interfaces of about 3,800 $Å^2$ in the ratcheted state and about 3,600 $Å^2$ in the tight state (Fig. 2b, c). The shelf–core interfaces in the two states are mostly overlapped (shown in green), but there are several contact points specific to either the ratcheted state or the tight state (shown in red and yellow, respectively). The central overlapped area of the interfaces is mainly hydrophobic, and the rotation axis of the ratcheting runs along it (Fig. 2b, c). The ratcheting axis forms an angle of about 50° to the floor of the channels on the core module (Fig. 2a). The core and shelf modules are connected by three peptide segments, the bridge helix, the loop consisting of β Tyr 998–Met 1005 (previously designated as 'switch 3')[10], and the loop consisting of β′ Ala 779–Ser 782 (designated hereafter as the 'hinge loop') (Fig. 2d, Supplementary Fig. 12). As the ratcheting axis runs close to β′ Pro 781 (*Escherichia coli* β′ Pro 502) within the hinge loop, the conformational change due to the ratcheting is negligibly small around this loop (Fig. 2e, f, Supplementary Fig. 12). By contrast, the bridge helix is much further from the ratcheting axis, as compared with the other two peptide segments (Fig. 2b, c), and the conformational change is significant, as described below in more detail.

On the other hand, the contact points specific to the ratcheted state are also distant from the ratcheting axis (Fig. 2b, c). In particular, an α-helix of the shelf module (β′ 685−696) contacts a β-hairpin from one of the α subunits (αB 184−191) (Fig. 2g). This contact probably limits the further ratcheting of the shelf module. Therefore, the ratcheted state is mostly at one extremity in the structural spectrum of bacterial RNAP, whereas the tight state observed in previous ECs[13,14] should be the other extremity.

The shelf module ratcheting results in a composite movement that expands the hybrid-binding site and shifts the shelf module forward, relative to the core module (Fig. 2e, f, Supplementary Text 6, Supplementary Fig. 13). The bridge helix exposes its central part (β′ 1084–1092), but buries its N-terminal part (β′ 1070–1083) and the carboxy-terminal part (β′ 1093–1102) within the core and shelf modules, respectively. As the two modules ratchet, the N- and C-terminal parts of the bridge helix shift relative to each other. Consequently, the conformation

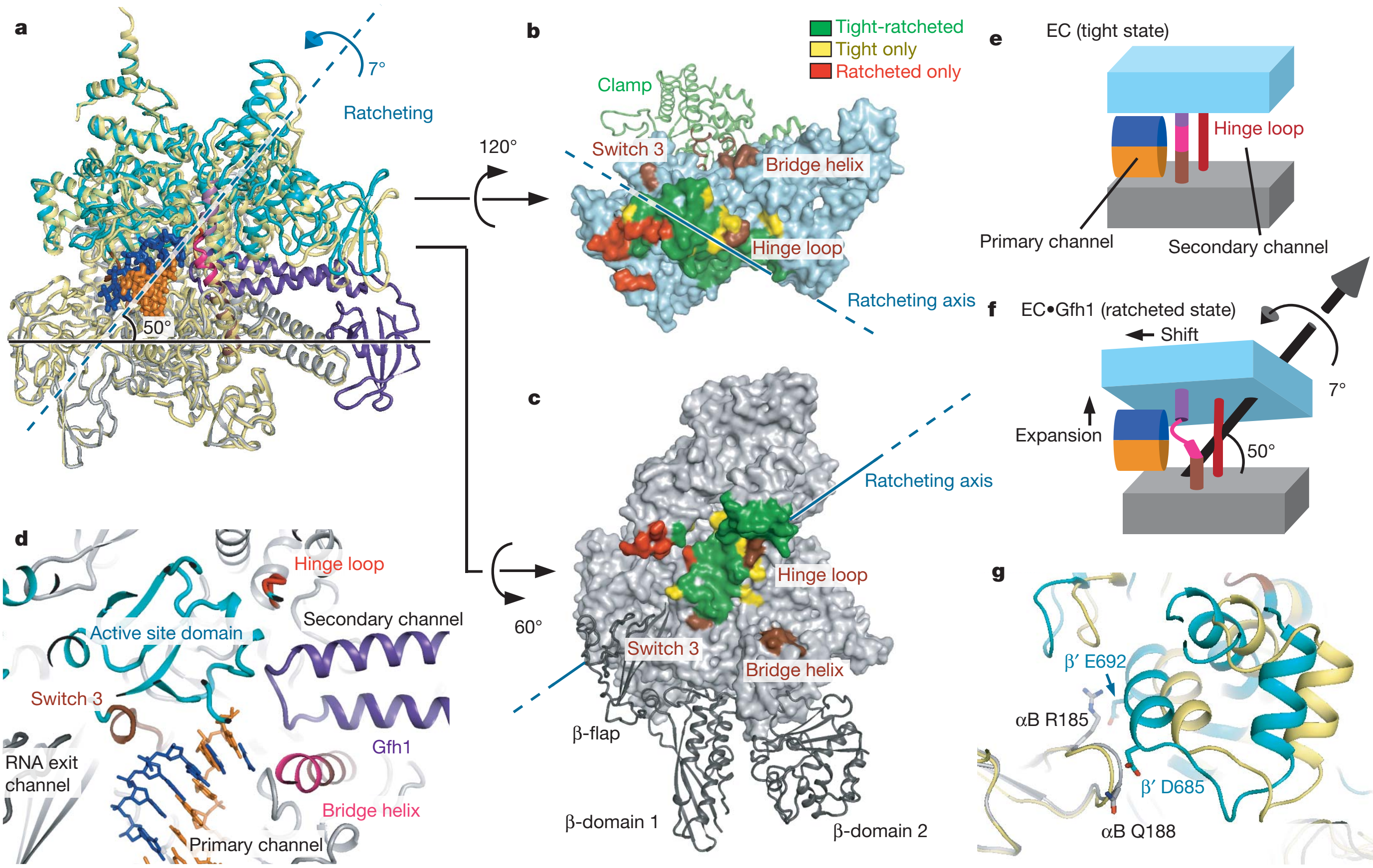


Figure 2 | Ratcheting of the shelf module. **a**, Superposition of EC•Gfh1 and EC (PDB 2O5I, yellow). The core modules of the two structures are superposed. The clamp, β′ NCD, and β domains are not shown. The dashed cyan line indicates the ratcheting axis. **b**, **c**, Interaction surface between the core and the shelf modules; two orientations are shown. **d**, The structures around the active site. The hinge loop is coloured red. **e**, **f**, A schematic drawing representing the structural differences between EC (**e**) and EC•Gfh1 (**f**). **g**, The contacts between the shelf and the core modules specific to the ratcheted state. EC•Gfh1 and EC are superposed by the core modules.

of β′ Thr 1088–Gly 1092 in the central part dramatically changes from the continuous α-helix in the previous ECs, and the two discontinuous α-helices are connected by the two non-helical residues, Ser 1091–Gly 1092 (Fig. 3a). It is intriguing that mutations of these two residues reportedly affect the RNAP activity[27]. On the other hand, the bridge helix is kinked in the RNAP structures of *T. aquaticus* core enzyme $(\alpha_2\beta\beta'\omega)$ and *T. thermophilus* holoenzyme $(\alpha_2\beta\beta'\omega\sigma)$ without nucleic acids[9,11]. However, the present kinked conformation is quite different from the previous ones (Supplementary Text 7 and Supplementary Fig. 14). Here, the two α-helices that are directly connected to the trigger loop are packed against the bridge-helix C-terminal region, and are therefore shifted together with it. To avoid steric hindrance with the tips of these two helices, the residues β′ Thr 1088, Ala 1089, Asp 1090 and Ser 1091 (*E. coli* β′ Thr 790, Ala 791, Asn 792 and Ser 793) protrude into the DNA•RNA hybrid binding site (Fig. 3b, Supplementary Text 6, Supplementary Fig. 13). Therefore, the conformational change of the bridge helix is likely to occur synchronously with the transition to the ratcheted state (Supplementary Text 7, 8). The direct interaction between Gfh1 and RNAP may affect the fine conformation of the bridge helix, by immobilizing its N-terminal region (see below).

Within the secondary channel, the Gfh1 NTD interacts tightly with parts of the shelf module (including the trigger loop) and the core module (including the secondary-channel coiled coil (β′ 958−1014) and the N-terminal part of the bridge helix), and fits particularly well with the narrowest region of the secondary channel (Fig. 3c, Supplementary Text 9, Supplementary Fig. 15, Supplementary Movie 1). The interaction between the Gfh1 NTD and the N-terminal part of the bridge helix seems to maintain the straight conformation of the N-terminal part, and to define the kinking point of the bridge helix in EC•Gfh1. It is impossible for the Gfh1 NTD to bind to the secondary channel in the tight EC in the same manner, as the channel is too narrow (Fig. 3d). Leu 33 of the Gfh1 NTD is located in the narrowest region of the secondary channel. A Gfh1 mutant with Leu 33 replaced by Trp (L33W) lacked transcription inhibition activity, probably because of an inability to bind (Fig. 3e, Supplementary Fig. 16). The bulky side chain of Trp seems to prevent the Gfh1 NTD from penetrating into the channel. This observation also indicates that the secondary channel cannot open beyond the width of the present ratcheted state. Consequently, Gfh1 just fits into the well-defined ratcheted state. Considering that Gfh1 cannot bind to RNAP in the tight state, because of steric hindrance, it is reasonable to postulate that Gfh1 traps a dynamically occurring, ratcheted state of RNAP. To examine this possibility, we performed crosslinking experiments. The results showed that an artificial disulphide bond or photo-crosslink was formed at the ratcheted state-specific interface between the core and shelf modules, even in the absence of Gfh1 (Supplementary Text 10, Supplementary Figs 17−19). Therefore, RNAP might spontaneously and dynamically alter its conformation, from the tight state to the ratcheted state, although the result could also be explained by more localized conformational fluctuations.

Owing to the interaction of the Gfh1 NTD with the narrowest part of the secondary channel, NTP entry would be prevented. Moreover, the presence of the Gfh1 NTD is compatible with the unfolded conformation of the trigger loop, but not with the NTP-induced, folded conformation[14,15] (Supplementary Fig. 20). The tip of the Gfh1 NTD is located near the RNAP active site (Supplementary Text 11, Supplementary Fig. 21). The end of the tip loop occupies the binding site of the β-γ phosphate groups in the NTP insertion step of the nucleotide

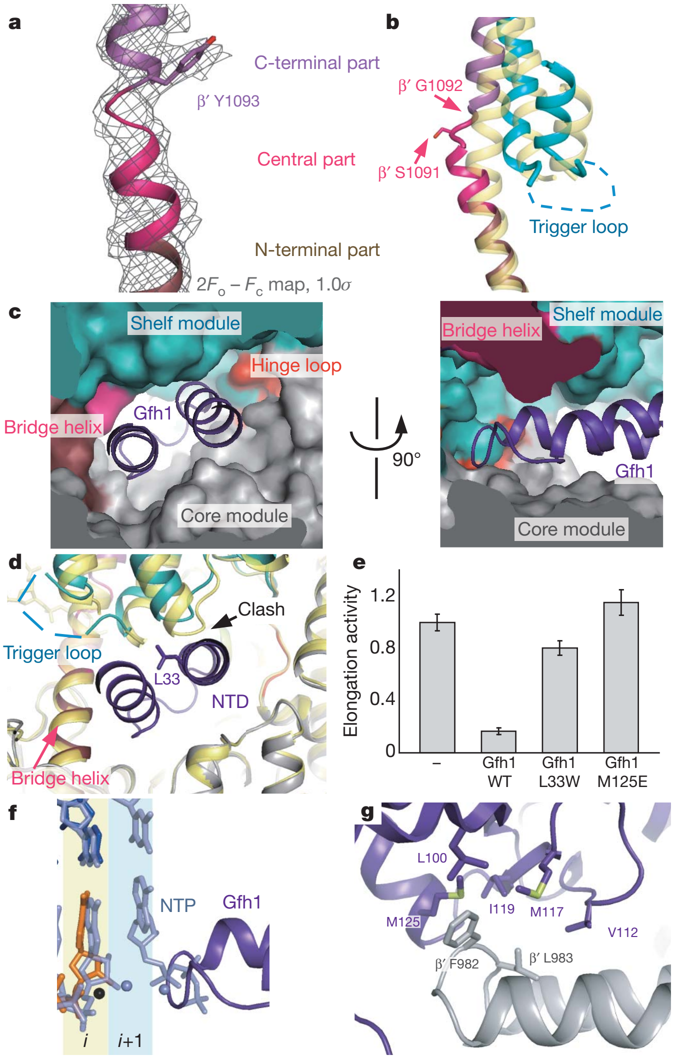


Figure 3 | The bridge helix and Gfh1. a, The structure and the electron density of the bridge helix in EC•Gfh1 (crystal 1, RNAP 1). **b**, The bridge helix and the helices neighbouring the trigger loop. *T. thermophilus* EC in the post-translocation state (PDB 2O5I) is superposed, and is coloured yellow. **c**, **d**, The Gfh1 NTD in the secondary channel. **d**, *T. thermophilus* EC in the post-translocation state (PDB 2O5I) is superposed. The black arrow indicates the position where Gfh1 would clash with EC in the tight state. **e**, Inhibition of the nucleotide addition reaction by Gfh1. The nucleotide addition reaction was performed by incubating the transcription elongation complex with 5 μM Gfh1 mutants. Error bars, s.d. ($n \geq 5$). **f**, A close-up view of the active site. *T. thermophilus* EC in the NTP-bound state (PDB 2O5J) is superposed, and is coloured light blue. *i* and *i* + 1 are the positions of the RNA 3′ end and the substrate NTP, respectively, in the nucleotide addition reaction. **g**, Interaction between the tip of the Gfh1 CTD and RNAP.

addition reaction (Fig. 3f). Furthermore, Gfh1 seems to stabilize the kinked bridge helix (Fig. 3b). An antibiotic, streptolydigin, also reportedly inhibited transcription by immobilizing the bridge helix in a fixed conformation[23]. Taken together, these observations provide the explanation for the inhibition of transcription elongation by Gfh1 (Supplementary Text 12, and Supplementary Figs 22, 23).

On the other hand, the C-terminal domain (CTD) of Gfh1 is bound with the edge of the secondary-channel coiled coil of the RNAP (Fig. 1a, b, Supplementary Text 13, Supplementary Fig. 24). The interaction involves the hydrophobic patch on the surface of the Gfh1 CTD (Fig. 3g). Therefore, we prepared a Gfh1protein with the M125E mutation within the hydrophobic patch, and observed a loss of inhibition activity (Fig. 3e). Therefore, the interaction between the hydrophobic patch of the Gfh1 CTD and the secondary-channel coiled coil is required for Gfh1 to bind to RNAP. Other Gre factors, such as GreA and GreB, share structural and sequence similarities with Gfh1 (refs 2–8). In particular, the presence of the hydrophobic patch on the CTD is well conserved. In fact, the M124E mutation of *E. coli* GreB (M125E of Gfh1) also reduces the transcript cleavage activity of GreB[3]. Therefore, the Gre factors seem to share a common interaction mode between the hydrophobic patch and the secondary-channel coiled coil, and probably bind to the ratcheted EC in a similar manner to Gfh1 (Supplementary Text 13, 14). Transcript cleavage stimulated by GreA and GreB would be performed in the ratcheted state. Although Pol II EC also changes its structure upon TFIIS binding[18,28], the observed change is much smaller than the transition to the ratcheted state of *T. thermophilus* EC upon Gfh1 binding (Supplementary Text 15).

The conformational changes of RNAP observed in the present structure, including the shelf module ratcheting and the clamp swinging, might have functional relevance to other stages of the transcription reaction, as the conformational changes should modulate the interactions of RNAP with nucleic acids. Therefore, we suggest that the conformational changes may play distinct roles in RNAP translocation and transcription termination (Supplementary Text 16, 17, Supplementary Figs 25−27). Experimental tests of these hypotheses will be required to assess the importance of these conformational changes in the absence of Gfh1.

METHODS SUMMARY

Structure determination. The structure of crystal 1 was solved by molecular replacement, using the coordinates of RNAP in *T. thermophilus* EC (PDB 2O5I)[13] as the search model. There are three RNAP molecules in the asymmetric unit, and each RNAP is bound with Gfh1. As the relative positions of the Gfh1 NTD and CTD differ from those in free Gfh1 (PDB 2F23)[7], they were separately placed in the electron density map. The model of the DNA•RNA hybrid was built in the extra electron density in the DNA•RNA hybrid channel. We further remodelled the coordinates of both the proteins and the nucleic acids with the program Coot[29]. Atomic positions and grouped *B*-factors were refined to 4.1 Å, by using the CNS program[30] (Supplementary Table 1). The refinement converged to R and R_{free} values of 26.2% and 31.8%, respectively; the latter was calculated from randomly-chosen 3% of reflections excluded from the refinement. The structure of crystal 2 was solved by molecular replacement, using the coordinates of the RNAP in crystal 1 as the search model. The models of Gfh1 and the DNA•RNA hybrid were built in the extra electron density. Refinement of the coordinates was performed to 4.3 Å with CNS. The final R and R_{free} values are 31.4% and 33.8%, respectively.

Transcription inhibition analysis. We prepared two mutant Gfh1 proteins (L33W and M125E). The elongation complex was reconstituted by incubating *T. thermophilus* RNAP with a nucleic acid scaffold containing template DNA, non-template DNA, and RNA. The RNA was 5′-radiolabelled using T4 polynucleotide kinase and [γ-^{32}P]-ATP. The nucleotide addition reaction was performed by incubating the transcription elongation complex in the presence of Gfh1 (the wild type or one of the mutants).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 October 2009; accepted 12 October 2010.
Published online 1 December 2010.

1. Borukhov, S., Polyakov, A., Nikiforov, V. & Goldfarb, A. GreA protein: a transcription elongation factor from *Escherichia coli. Proc. Natl Acad. Sci. USA* **89,** 8899–8902 (1992).
2. Stebbins, C. E. *et al.* Crystal structure of the GreA transcript cleavage factor from *Escherichia coli. Nature* **373,** 636–640 (1995).
3. Vassylyeva, M. N. *et al.* The carboxy-terminal coiled-coil of the RNA polymerase β′-subunit is the main binding site for Gre factors. *EMBO Rep.* **8,** 1038–1043 (2007).
4. Hogan, B. P., Hartsch, T. & Erie, D. A. Transcript cleavage by *Thermus thermophilus* RNA polymerase. Effects of GreA and anti-GreA factors. *J. Biol. Chem.* **277,** 967–975 (2002).
5. Laptenko, O. & Borukhov, S. Biochemical assays of Gre factors of *Thermus thermophilus. Methods Enzymol.* **371,** 219–232 (2003).
6. Lamour, V., Hogan, B. P., Erie, D. A. & Darst, S. A. Crystal structure of *Thermus aquaticus* Gfh1, a Gre-factor paralog that inhibits rather than stimulates transcript cleavage. *J. Mol. Biol.* **356,** 179–188 (2006).
7. Laptenko, O. *et al.* pH-dependent conformational switch activates the inhibitor of transcription elongation. *EMBO J.* **25,** 2131–2141 (2006).

8. Symersky, J. *et al.* Regulation through the RNA polymerase secondary channel. Structural and functional variability of the coiled-coil transcription factors. *J. Biol. Chem.* **281,** 1309–1312 (2006).
9. Zhang, G. *et al.* Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98,** 811–824 (1999).
10. Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292,** 1863–1876 (2001).
11. Vassylyev, D. G. *et al.* Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417,** 712–719 (2002).
12. Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292,** 1876–1882 (2001).
13. Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. H. & Artsimovitch, I. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448,** 157–162 (2007).
14. Vassylyev, D. G. *et al.* Structural basis for substrate loading in bacterial RNA polymerase. *Nature* **448,** 163–168 (2007).
15. Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D. & Kornberg, R. D. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127,** 941–954 (2006).
16. Westover, K. D., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* **303,** 1014–1016 (2004).
17. Westover, K. D., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* **119,** 481–489 (2004).
18. Kettenberger, H., Armache, K. J. & Cramer, P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell* **16,** 955–965 (2004).
19. Brueckner, F. & Cramer, P. Structural basis of transcription inhibition by α-amanitin and implications for RNA polymerase II translocation. *Nature Struct. Mol. Biol.* **15,** 811–818 (2008).
20. Sydow, J. F. *et al.* Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol. Cell* **34,** 710–721 (2009).
21. Wang, D. *et al.* Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science* **324,** 1203–1206 (2009).
22. Temiakov, D. *et al.* Structural basis of transcription inhibition by antibiotic streptolydigin. *Mol. Cell* **19,** 655–666 (2005).
23. Tuske, S. *et al.* Inhibition of bacterial RNA polymerase by streptolydigin: stabilization of a straight-bridge-helix active-center conformation. *Cell* **122,** 541–552 (2005).
24. Tagami, S., Sekine, S., Kumarevel, T., Yamamoto, M. & Yokoyama, S. Crystallization and preliminary X-ray crystallographic analysis of *Thermus thermophilus* transcription elongation complex bound to Gfh1. *Acta Crystallogr. F* **66,** 64–68 (2010).
25. Darst, S. A. *et al.* Conformational flexibility of bacterial RNA polymerase. *Proc. Natl Acad. Sci. USA* **99,** 4296–4301 (2002).
26. Murakami, K. S., Masuda, S. & Darst, S. A. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* **296,** 1280–1284 (2002).
27. Tan, L., Wiesler, S., Trzaska, D., Carney, H. C. & Weinzierl, R. O. Bridge helix and trigger loop perturbations generate superactive RNA polymerases. *J. Biol.* **7,** 40 (2008).
28. Kettenberger, H., Armache, K. J. & Cramer, P. Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* **114,** 347–357 (2003).
29. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
30. Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nature Protocols* **2,** 2728–2733 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is based on experiments performed at SPring-8 (with the approval of the Japan Synchrotron Radiation Research Institute) and at the Swiss Light Source (SLS). We thank N. Shimizu for supporting our data collection at SPring-8 beamline BL41XU; T. Tomizaki and C. Schulze-Briese for supporting our data collection at SLS beamline X06SA; and Y. Fujii for assisting with our data collection and for comments. We thank T. Tanaka and K. Sakamoto for assistance in protein preparation. This work was supported in part by a Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Young Scientists (to S.-i.S.), a JSPS Grant-in-Aid for Scientific Research (to S.i.-S. and S.Y.), and the Targeted Proteins Research Program (TPRP), the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. S.T. was supported by the JSPS Global Centers of Excellence Program (Integrative Life Science Based on the Study of Biosignaling Mechanisms).

Author Contributions S.T., S.-i.S., T. K. and S.Y. designed the research. S.T. and S.-i.S. performed the structural analysis. M.Y. supported the structural analysis. S.T., S.-i.S., N.H., S.K. and K.S. performed the disulphide-bonding and/or photo-crosslinking analyses. S.T. and Y.M. performed the biochemical analysis of Gre factors. S.-i.S. created the movies. S.T., S.-i.S. and S.Y. wrote the paper.

Author Information The structures of EC•Gfh1 have been deposited in the Protein Data Bank, under accession numbers 3AOH (crystal 1) and 3AOI (crystal 2). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.-i.S. (sekine@biochem.s.u-tokyo.ac.jp) or S.Y. (yokoyama@biochem.s.u-tokyo.ac.jp).

METHODS

Structure determination. Crystallization and data collection were described previously[24]. The data were reprocessed with the XDS program[31]. The structure for crystal 1 was solved by molecular replacement with the program Phaser[32], using the coordinates of the core enzyme portion of *T. thermophilus* EC (PDB 2O5I)[13] as the search model. The asymmetric unit contains three RNAP molecules. Each RNAP was divided into 25−26 rigid bodies, and their positions were refined with the program CNS version 1.2[30,33]. Several of the rigid bodies deviated substantially from the electron density, and they were manually adjusted to the density with the program Coot[29]. For the tip portion of the β′ non-conserved domain (β′ NCD, β′ 132−454), the coordinates of the *T. thermophilus* holoenzyme (PDB 3DXJ)[34] were used. Several rounds of rigid body refinement and manual adjustment were performed. In each RNAP molecule, extra electron density, corresponding to Gfh1, was observed. The NTD and CTD coordinates of free Gfh1 (PDB 2F23)[7] were separately placed in the $2F_o - F_c$ electron density map. One of the RNAP molecules in the asymmetric unit exhibited extra electron density corresponding to the DNA•RNA hybrid in the DNA•RNA hybrid binding site, for which we built the hybrid model. The coordinates of the DNA•RNA hybrid in *S. cerevisiae* EC (PDB 2VUM)[19] were used as the starting model. The electron density for the nucleic acids in the other two RNAP complexes was weak, probably owing to low occupancy and/or high mobility, and therefore, we did not build their models.

The structures of the N- and C-terminal parts of the bridge helix in EC•Gfh1 are similar to those in the previous EC (2O5I)[13], while the central part of the bridge helix in the present complex exhibits a conformational change, due to the ratcheting of the core and shelf modules. The region of β′1086–1090 assumes a helical, but slightly curved conformation, and the model was built by adjusting the corresponding region in the previous EC (2O5I) to the electron density. Most parts of the bridge helix (β′ 1070−1090 and β′ 1093−1102) maintained the helical conformation. β′ Ser 1091 and β′ Gly 1092 were placed to link the two discontinuous helices, while fitting their main chains into the electron density. For this rebuilding, the position of β′ Tyr 1093, which was identified by the electron density of its large side chain, was helpful (Fig. 3a). The coordinates of both the proteins and the nucleic acids were further refined with the program Coot. The atomic positions and the grouped *B*-factors were refined to 4.1 Å, by using CNS with strong NCS restraints among the three complexes in the asymmetric unit (Supplementary Table 1). Refinement was monitored by R_{free}, calculated from 3% of the reflections that were excluded from the refinement.

The structure for crystal 2 was solved by molecular replacement, using the coordinates of RNAP and Gfh1 in crystal 1 as the search model. A model of the DNA•RNA hybrid was placed in each RNAP in the electron density map, and positional refinement of the coordinates was performed to 4.3 Å with CNS. For the calculation of R_{free}, the same reflections as those chosen for crystal 1 were used.

The rigid bodies used in the structural refinement allowed us to define the mobile modules of *T. thermophilus* RNAP. We first superposed the RNAPs in the present EC•Gfh1 and the previous EC (2O5I)[13] by certain rigid bodies, and then inspected the rigid bodies that superposed well concurrently. The masses of the co-relocated rigid bodies were defined as modules. Finally, we confirmed that the defined mobile modules relocated separately by the RigidFinder program[35].

Disulphide-bonding assay for the ratcheted RNAP. We constructed a plasmid that allows co-expression of the α, β, β′ and ω subunits of *T. thermophilus* RNAP in *E. coli*, for the preparation of recombinant *T. thermophilus* RNAP (pRpoBCAZ, to be published elsewhere). The recombinant RNAPs (the wild type and the α Q188C-β′ D685C mutant) were expressed using this system. They were purified by the procedure used for the natural RNAP core enzyme from *T. thermophilus* cells[36], except that the cell lysate was heat-treated at 70 °C for 30 min, in order to denature most of the non-thermophilic *E. coli* proteins. Then, the RNAPs were fractionated by polyethyleneimine precipitation, followed by ammonium sulphate precipitation. The recombinant RNAPs were further purified by chromatography on Q-Sepharose and Superdex pg200 columns (GE Healthcare Biosciences).

For the disulphide-bonding analysis, the recombinant RNAP (the wild type or the α Q188C-β′ D685C mutant) was dissolved in 75 mM Tris-HCl buffer (pH 8.1), containing 50 mM KCl, 10 mM $MgCl_2$ and 1 mM DTT. Each of the RNAPs (0.2 μM) was incubated with 0.7 μM of the nucleic acid scaffold (DNATS/DNANT/RNA14) or 10 μM of *T. thermophilus* Gfh1 for 30 min. Then, 2.5 mM glutathione disulphide (GSSG) was added to each RNAP solution for the mild oxidation of Cys residues. The mixtures were analysed by SDS–PAGE, using sample buffer lacking a reducing agent. The formation of a disulphide bond between α C188 and β′ C685 was confirmed by the appearance of an extra band with low mobility, which corresponded to the crosslinked α and β′ subunits (Supplementary Fig. 17). The sequences of the nucleic acids are as follows. The RNA oligomer: RNA14, UUUUUGAGUCUGCGGCGAU. The DNA oligomers: DNATS, AACATACGGCTCGGACAGAGGTCCTGTCTGAATCGATATCGC CGC; DNANT, CGATTCAGACAGGACCTCTGTCCGAGCCGTATGTT. The nucleic acid scaffold was designed by modifying the previously reported EC14 scaffold, which forms a stable elongation complex with *T. thermophilus* RNAP[37].

Photo-crosslinking assay of the ratcheted RNAP. *p*-Benzoyl-L-phenylalanine (*p*Bpa) is a photo-crosslinker that can be position-specifically incorporated into a recombinant protein[38,39]. The gene encoding a *p*Bpa-specific variant of *Methanococcus jannaschii* tyrosyl-tRNA synthetase[38], under the control of the *E. coli tyrS* promoter, was cloned in the pACYC184 vector, together with three copies of the *M. jannaschii* amber suppressor tRNA gene[39], to create a vector for the expression of the *p*Bpa-specific tRNA synthetase and tRNA (ppbpaRS-3MJR1). Each tRNA gene had an *E. coli lpp* promoter and an *rrnC* terminator. The artificial operons for overproducing minor tRNA species, including the minor $tRNA^{Pro}$, were described previously[40], and were cloned in a kanamycin-resistant plasmid carrying the CloDF13-derived replication origin, to create pMINOR2.

The *rpoA* gene, C-terminally tagged with FLAG, was engineered, using a QuikChange mutagenesis kit (Stratagene), to have an amber codon in place of Arg 185, for producing the RNAP α-subunit with Arg 185 replaced with *p*Bpa (*p*Bpa 185) (ref. 41). The *rpoC* gene was engineered to have a methionine codon in place of Glu 692, for producing the β′ subunit with the E692M substitution. The *rpoA* and *rpoC* genes in pRpoBCAZ were replaced by these mutant genes, and the vector was introduced into BL21 Star(DE3) cells (Invitrogen) harbouring the ppbpaRS-3MJR1 and pMINOR2 plasmids. The cells were grown in LB medium containing 1 mM *p*Bpa, and the gene expression was induced by the addition of 1 mM IPTG at the mid-log phase. After a further 4-h incubation, the cells were harvested and lysed by sonication in buffer A (40 mM Tris-HCl (pH 7.7), 500 mM NaCl, 10 mM EDTA, 10 mM 2-mercaptoethanol, 5% glycerol, and Complete protease inhibitor cocktail tablets (Roche)). The wild-type and engineered RNAP core enzymes were roughly purified by heat-treatment. For the photo-crosslinking, the proteins were exposed to light at 365 nm for 30 min on ice[41] in a 24-well cell culture plate (BD Biosciences), followed by SDS–PAGE, Coomassie brilliant blue staining, and western blotting with an anti-FLAG antibody (Sigma) (Supplementary Fig. 19a, 19b). The RNAP core enzymes were further purified as described above. Then, 2 μM RNAP (wild type or mutant) was incubated with the nucleic acid scaffold (2.5 μM) or Gfh1 (10 μM) for 30 min at room temperature in 50 mM HEPES-NaOH buffer (pH 7.5), containing 50 mM KCl, 10 mM $MgCl_2$ and 1 mM DTT, followed by the photo-crosslinking step (Supplementary Fig. 19c).

Transcription inhibition analysis. In the previous study, we constructed a plasmid for the expression of wild-type Gfh1 in *E. coli*[24]. The expression plasmids for Gfh1 variants (L33W and M125E) were generated by introducing mutations to the plasmid encoding wild-type Gfh1. In addition, we constructed plasmids for the expression of wild type and mutant *T. thermophilus* GreA in *E. coli*. The wild-type and mutant Gre proteins were expressed as described previously[24], and were then purified by chromatography on Toyopearl Super-Q and Butyl columns (Tosoh Bioscience). The transcription elongation complex was reconstituted by incubating 0.1 μM of *T. thermophilus* RNAP with 0.1 μM of the nucleic acid scaffold (DNATS/DNANT/RNA14 or RNA15) for 30 min, where the sequence of RNA15 is UUUUUG AGUCUGCGGCGAUA. The RNA was 5′-radiolabelled using T4 polynucleotide kinase and [γ-^{32}P]-ATP. The nucleotide addition reaction was performed by incubating the transcription elongation complex of RNA14 with 5 μM Gfh1 (wild type or a mutant) at 20 °C in 50 mM MES-NaOH buffer (pH 6.5), containing 50 mM KCl, 10 mM $MgCl_2$, 1 mM DTT and 20 μM ATP and that of RNA15 with 2.5 μM Gfh1 (wild type or a mutant) at 55 °C in 50 mM MES-NaOH buffer (pH 6.5), containing 50 mM KCl, 1 mM $MgCl_2$, 1 mM DTT and 20 μM UTP. The RNA was analysed by denaturing (8 M urea) PAGE.

31. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26,** 795–800 (1993).
32. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
33. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
34. Mukhopadhyay, J. *et al.* The RNA polymerase "switch region" is a target for inhibitors. *Cell* **135,** 295–307 (2008).
35. Abyzov, A., Bjornson, R., Felipe, M. & Gerstein, M. RigidFinder: a fast and sensitive method to detect rigid blocks in large macromolecular complexes. *Proteins* **78,** 309–324 (2010).
36. Vassylyeva, M. N. *et al.* Purification, crystallization and initial crystallographic analysis of RNA polymerase holoenzyme from *Thermus thermophilus*. *Acta Crystallogr. D* **58,** 1497–1500 (2002).
37. Kashkina, E. *et al.* Elongation complexes of *Thermus thermophilus* RNA polymerase that possess distinct translocation conformations. *Nucleic Acids Res.* **34,** 4036–4045 (2006).

38. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **99,** 11020–11024 (2002).
39. Sakamoto, K. *et al.* Genetic encoding of 3-iodo-L-tyrosine in *Escherichia coli* for single-wavelength anomalous dispersion phasing in protein crystallography. *Structure* **17,** 335–344 (2009).
40. Chumpolkulwong, N. *et al.* Translation of 'rare' codons in a cell-free protein synthesis system from *Escherichia coli*. *J. Struct. Funct. Genomics* **7,** 31–36 (2006).
41. Hino, N. *et al.* Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nature Methods* **2,** 201–206 (2005).

doi:10.1038/nature09578

A substantial population of low-mass stars in luminous elliptical galaxies

Pieter G. van Dokkum[1] & Charlie Conroy[2,3]

The stellar initial mass function (IMF) describes the mass distribution of stars at the time of their formation and is of fundamental importance for many areas of astrophysics. The IMF is reasonably well constrained in the disk of the Milky Way[1] but we have very little direct information on the form of the IMF in other galaxies and at earlier cosmic epochs. Here we report observations of the Na I doublet[2,3] and the Wing–Ford molecular FeH band[4,5] in the spectra of elliptical galaxies. These lines are strong in stars with masses less than $0.3M_\odot$ (where $M_\odot$ is the mass of the Sun) and are weak or absent in all other types of stars[5–7]. We unambiguously detect both signatures, consistent with previous studies[8] that were based on data of lower signal-to-noise ratio. The direct detection of the light of low-mass stars implies that they are very abundant in elliptical galaxies, making up over 80% of the total number of stars and contributing more than 60% of the total stellar mass. We infer that the IMF in massive star-forming galaxies in the early Universe produced many more low-mass stars than the IMF in the Milky Way disk, and was probably slightly steeper than the Salpeter form[9] in the mass range $0.1M_\odot$ to $1M_\odot$.

We obtained spectra of eight of the most luminous and massive galaxies in the nearby Universe: four of the brightest early-type galaxies in the Virgo cluster and four in the Coma cluster. The galaxies were selected to have velocity dispersions $\sigma > 250\,\mathrm{km\,s^{-1}}$, and were observed with the Low-Resolution Imaging Spectrometer[10] (LRIS) on the Keck I telescope. In 2009 the red arm of LRIS was outfitted with fully depleted charge-coupled devices (CCDs), which have excellent sensitivity out to wavelengths of $\lambda > 9{,}000$ Å and almost no fringing. The individual spectra of the four galaxies in each of the two clusters were de-redshifted, averaged and binned to a resolution of 8 Å.

In Fig. lb and c we show the spectral region near the $\lambda = 8{,}183$, $\lambda = 8{,}195$ Na I doublet for the Virgo and Coma galaxies. The doublet appears as a single absorption feature due to Doppler broadening. In Fig. le and f we show the region around the $\lambda = 9{,}916$ Wing–Ford band for the Virgo galaxies. This region could not be observed with sufficient signal in Coma because it is redshifted to 1.015 µm. The spectra are of very high quality. The median 1σ scatter of the four galaxies around the average spectrum is only about 0.3% per spectral bin. The median absolute difference between the Virgo and Coma spectra is 0.4% per spectral bin.

Both the Na I doublet and the Wing–Ford band are unambiguously detected. The central wavelength of the observed Na I line coincides with the weighted average wavelength of the doublet and the observed Wing–Ford band has the characteristic asymmetric profile reflecting the $A^4\Delta$ to $X^4\Delta$ transition of FeH (ref. 5). The Na I index is 0.058 ± 0.006 mag in the Virgo galaxies and 0.057 ± 0.007 mag in Coma. The Wing–Ford index in Virgo galaxies is 0.027 ± 0.005. The uncertainties are determined from the scatter among the individual galaxies. We note that any residual systematic problems with the detector or atmosphere are incorporated in this scatter, as the features were originally redshifted to a different observed wavelength range for each of the galaxies.

The immediate implication is that stars with masses less than $0.3M_\odot$ are present in substantial numbers in the central regions of elliptical galaxies. Such low-mass stars are impossible to detect individually in external galaxies, because they are too faint: Barnard's star would have a K-band magnitude of 39 at the distance of the Virgo cluster. This in turn implies that there was a channel for forming low-mass stars in the progenitors of luminous early-type galaxies in clusters. These star-forming progenitors are thought to be relatively compact galaxies at $z = 2 - 5$ with star-formation rates of tens or hundreds of solar masses per year. Some studies have suggested truncated IMFs for such galaxies[11], with a cut-off below $1M_\odot$. Such dwarfless IMFs are effectively ruled out by the detection of the Na I lines and the Wing–Ford band.

We turn to stellar population synthesis models to quantify the number of low-mass stars in elliptical galaxies. As discussed in detail in the Supplementary Information, we use a flexible stellar population synthesis code[12] combined with an extensive empirical library of stellar near-infrared spectra[13]. In Fig. 1 we show synthetic spectra in both spectral regions for different choices of the IMF[1,9,14], including IMFs that are steeper than the Salpeter form. Away from the Na I doublet and the Wing-Ford band all models fit very well, with differences between data and the best-fitting model of less than 0.5% over the entire spectral range. Predicted Na I and Wing-Ford line indices are compared to the observed values in Fig. 2. The data prefer IMFs with substantial dwarf populations. The best fits are obtained for a logarithmic IMF slope of $x \approx -3$, a more dwarf-rich ('bottom-heavy') IMF than even the Salpeter form, which has $x = -2.35$. A Kroupa IMF (which is appropriate for the Milky Way) is inconsistent with the Wing–Ford data at $>2\sigma$ and inconsistent with the Na I data at $>4\sigma$, as are IMFs with even more suppressed dwarf populations[14–16]. We note that the $x = -3$ IMF also provides a much better fit to the region around 0.845 µm than any of the other forms. Taking the Salpeter IMF as a limiting case, we find that stars with masses of $0.1M_\odot$ to $0.3M_\odot$ make up at least 80% of the total number of living stars in elliptical galaxies, and contribute at least 60% of the total stellar mass.

Although the formal uncertainty in the derived IMF slope is small we stress that some unknown systematic effect could be present in the stellar population synthesis modelling. In particular, weak features in the spectra of giant stars in elliptical galaxies may be incorrectly represented by the Milky Way giants that we use. It may also be that the Na abundance of low-mass stars in elliptical galaxies is different from that of low-mass stars in the Milky Way. The fact that all models fit the spectra of both the Virgo and Coma galaxies extremely well outside of the IMF-sensitive regions gives some confidence in our approach; as we show in the Supplementary Information, the quality of the fit constrains possible contamination of spectral features such as TiO lines.

Besides model uncertainties, the interpretation may be complicated by the fact that we are constraining the IMF some ten billion years after the stars were formed. It is now generally thought that elliptical galaxies have undergone several (or many) mergers with other galaxies after their initial collapse[17], which may imply that the stellar population is

[1]Astronomy Department, Yale University, New Haven, Connectitcut, USA. [2]Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey, USA. [3]Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts, USA.

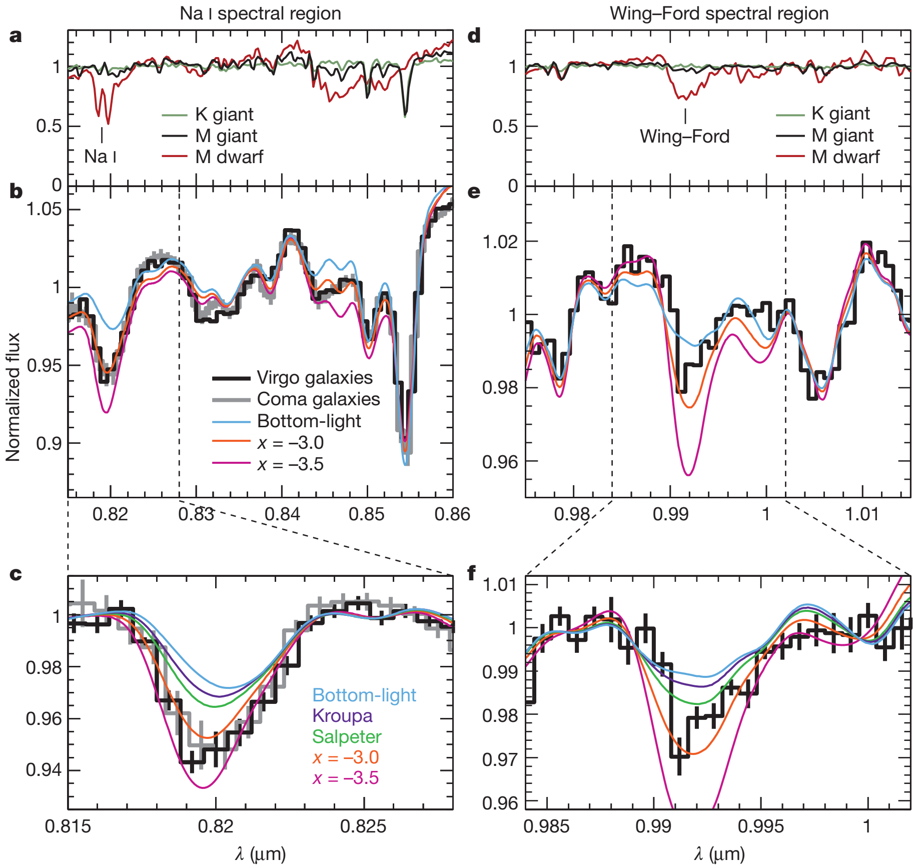


Figure 1 | Detection of the Na I doublet and the Wing–Ford band. **a**, Spectra in the vicinity of the $\lambda = 8{,}183$, $\lambda = 8{,}195$ Na I doublet for three stars from the IRTF library[12]: a K0 giant, which dominates the light of old stellar populations; an M6 dwarf, the (small) contribution of which to the integrated light is sensitive to the form of the IMF at low masses; and an M3 giant, which has potentially contaminating TiO spectral features in this wavelength range. **b**, Averaged Keck/LRIS spectra of NGC 4261, NGC 4374, NGC 4472 and NGC 4649 in the Virgo cluster (black line) and NGC 4840, NGC 4926, IC 3976 and NGC 4889 in the Coma cluster (grey line). Four exposures of 180 s were obtained for each galaxy. The one-dimensional spectra were extracted from the reduced two-dimensional data by summing the central 4″, which corresponds to about 0.4 kpc at the distance of Virgo and about 1.8 kpc at the distance of Coma. We found little or no dependence of the results on the choice of aperture. Coloured lines show stellar population synthesis models for a dwarf-deficient 'bottom-light' IMF[14], a dwarf-rich 'bottom-heavy' IMF with $x = -3$, and an even more dwarf-rich IMF. The models are for an age of 10 Gyr and were smoothed to the average velocity dispersion of the galaxies. The $x = -3$ IMF fits the spectrum remarkably well. **c**, Spectra and models around the dwarf-sensitive Na I doublet. A Kroupa IMF, which is appropriate for the Milky Way, does not produce a sufficient number of low-mass stars to explain the strength of the absorption. An IMF steeper than Salpeter appears to be needed. **d**–**f**, Spectra and models near the $\lambda = 9{,}916$ Wing–Ford band. The observed Wing–Ford band also favours an IMF that is more abundant in low-mass stars than the Salpeter IMF. All spectra and models were normalized by fitting low-order polynomials (excluding the feature of interest). The polynomials were quadratic in **a**, **b**, **d** and **e** and linear in **c** and **f**.

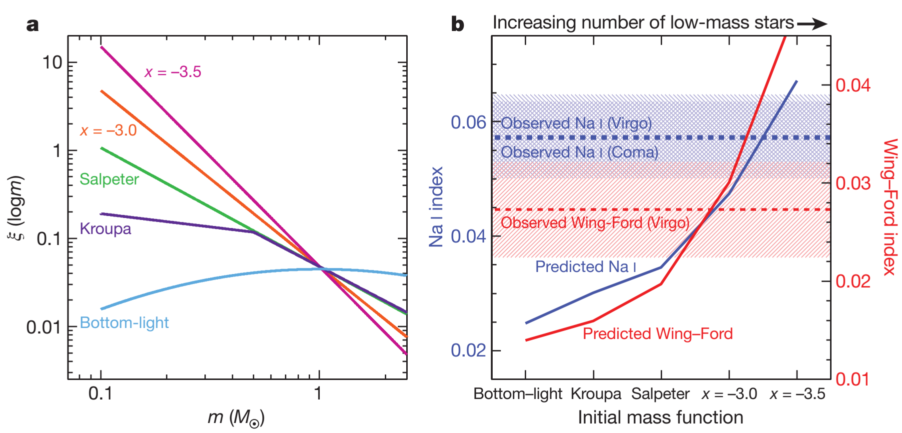


Figure 2 | Constraining the IMF. **a**, Various stellar IMFs, ranging from a 'bottom-light' IMF with strongly suppressed dwarf formation[14] (light blue) to an extremely 'bottom-heavy' IMF with a slope $x = -3.5$. The IMFs are normalized at $1M_{\odot}$, because stars of approximately one solar mass dominate the light of elliptical galaxies. **b**, Comparison of predicted line Na I and Wing–Ford indices with the observed values. The indices were defined to be analogous to those in refs 4 and 8. The Na I index has central wavelength 0.8195 μm and side bands at 0.816 μm and 0.825 μm. The Wing–Ford index has central wavelength 0.992 μm and side bands at 0.985 μm and 0.998 μm. The central bands and side bands are all 20 Å wide. Both observed line indices are much stronger than expected for a Kroupa IMF. The best fits are obtained for IMFs that are slightly steeper than Salpeter.

more complex than our single-age, single-metallicity model. On the other hand, these accretion events probably mostly add stars at large radii[18] and may not have affected the core regions very much. It will be interesting to search for gradients in dwarf-sensitive features with more extensive data[19].

Our results are consistent with previous studies of the near-infrared spectra of elliptical galaxies[8,20,21]. They are also consistent with recent dynamical and lensing constraints on the IMF in elliptical galaxies with large velocity dispersions[22] and directly identify the stars that are responsible for their high masses: the dynamical data cannot distinguish dwarf-rich IMFs from dwarf-deficient IMFs because the latter have a large amount of mass in stellar remnants[14]. A steep IMF for elliptical galaxies is also qualitatively consistent with the apparently higher number of low-mass stars in the Milky Way bulge than in the disk[23]. Our best-fitting IMF does not appear to be consistent with the observed colour and M/L evolution of massive cluster galaxies[14], which suggest an IMF with a slope $x \approx -1$ around $1M_{\odot}$. Interpreting the evolution of the colours and luminosities of elliptical galaxies relies on the assumption that these galaxies evolve in a self-similar way, which may not be valid[18,24]. It could also be that the form of the IMF is more complex than a power law.

Our results also seem inconsistent with theoretical arguments for dwarf-deficient IMFs at high redshift, which have centred on the idea that the characteristic mass of stars scales with the Jeans mass in molecular clouds[25,26]. The Jeans mass has a strong temperature dependence and it has been argued that relatively high ambient temperatures in high-redshift star-forming galaxies may have set a lower boundary to the characteristic mass in the progenitors of elliptical galaxies[14,25]. However, the Jeans mass also scales with density, and the gas densities in the star-forming progenitors of the cores of elliptical galaxies were almost certainly significantly higher than typical densities of star-forming regions in the Milky Way. Numerical simulations suggest that the formation of low-mass stars becomes inevitable if sufficiently high densities are reached on sub-parsec scales[27]. Furthermore, recent semi-analytic models of the thermal evolution of gas clouds have emphasized the effects of dust-induced cooling[28], which is relatively insensitive to the ambient temperature and particularly effective at high densities. Timescale arguments suggest that the physical conditions expected in starburst galaxies at high redshift might even enhance low-mass star formation, rather than suppress it[29].

Taken at face value, our results imply that the form of the IMF is not universal but depends on the prevailing physical conditions: Kroupa-like in quiet, star-forming disks and dwarf-rich in the progenitors of massive elliptical galaxies. This informs models of star formation and has important implications for the interpretation of observations of galaxies in the early Universe. The stellar masses and star-formation rates of distant galaxies are usually estimated from their luminosities, assuming some form of the IMF[30]. Our results suggest that a different form should be used for different galaxies, greatly complicating the analysis. The bottom-heavy IMF advocated here may also require a relatively low fraction of dark matter within the central regions of nearby massive galaxies[22].

Received 7 July; accepted 12 October 2010.
Published online 1 December 2010.

1. Kroupa, P. On the variation of the initial mass function. *Mon. Not. R. Astron. Soc.* **322,** 231–246 (2001).
2. Faber, S. M. & French, H. B. Possible M dwarf enrichment in the semistellar nucleus of M31. *Astrophys. J.* **235,** 405–412 (1980).
3. Schiavon, R. P., Barbuy, B., Rossi, S. C. F. & Milone, A. The near-infrared Na I doublet feature in M stars. *Astrophys. J.* **479,** 902–908 (1997).
4. Wing, R. F. & Ford, W. K. The infrared spectrum of the cool dwarf Wolf 359. *Publ. Astron. Soc. Pacif.* **81,** 527–529 (1969).
5. Schiavon, R. P., Barbuy, B. & Singh, P. D. The FeH Wing-Ford band in spectra of M stars. *Astrophys. J.* **484,** 499–510 (1997).
6. Schiavon, R. P., Barbuy, B. & Bruzual, A. G. Near-infrared spectral features in single-aged stellar populations. *Astrophys. J.* **532,** 453–460 (2000).
7. Cushing, M. C., Rayner, J. T., Davis, S. P. & Vacca, W. D. FeH absorption in the near-infrared spectra of late M and L dwarfs. *Astrophys. J.* **582,** 1066–1072 (2003).
8. Couture, J. & Hardy, E. The low-mass stellar content of galaxies—constraints through hybrid population synthesis near 1 micron. *Astrophys. J.* **406,** 142–157 (1993).
9. Salpeter, E. E. The luminosity function and stellar evolution. *Astrophys. J.* **121,** 161–167 (1955).
10. Oke, J. B. *et al.* The Keck low-resolution imaging spectrometer. *Publ. Astron. Soc. Pacif.* **107,** 375–385 (1995).
11. Baugh, C. M. *et al.* Can the faint submillimetre galaxies be explained in the Λ cold dark matter model? *Mon. Not. R. Astron. Soc.* **356,** 1191–1200 (2005).
12. Conroy, C., Gunn, J. E. & White, M. The propagation of uncertainties in stellar population synthesis modeling. I. The relevance of uncertain aspects of stellar evolution and the initial mass function to the derived physical properties of galaxies. *Astrophys. J.* **699,** 486–506 (2009).
13. Rayner, J. T., Cushing, M. C. & Vacca, W. D. The Infrared Telescope Facility (IRTF) spectral library: cool stars. *Astrophys. J.*, Suppl. 623, 289–432 (2009).
14. van Dokkum, P. G. Evidence of cosmic evolution of the stellar initial mass function. *Astrophys. J.* **674,** 29–50 (2008).
15. Fardal, M. A., Katz, N., Weinberg, D. H. & Davé, R. On the evolutionary history of stars and their fossil mass and light. *Mon. Not. R. Astron. Soc.* **379,** 985–1002 (2007).
16. Davé, R. The galaxy stellar mass-star formation rate relation: evidence for an evolving stellar initial mass function? *Mon. Not. R. Astron. Soc.* **385,** 147–160 (2008).
17. Naab, T., Johansson, P. H., Ostriker, J. P. & Efstathiou, G. Formation of early-type galaxies from cosmological initial conditions. *Astrophys. J.* **658,** 710–720 (2007).
18. van Dokkum, P. G. *et al.* The growth of massive galaxies since $z = 2$. *Astrophys. J.* **709,** 1018–1041 (2010).
19. Boroson, T. A. & Thompson, I. B. Color distributions in early type galaxies. III – Radial gradients in spectral features. *Astron. J.* **101,** 111–126 (1991).
20. Cenarro, A. J., Gorgas, J., Vazdekis, A., Cardiel, N. & Peletier, R. F. Near-infrared line-strengths in elliptical galaxies: evidence for initial mass function variations? *Mon. Not. R. Astron. Soc.* **339,** L12–L16 (2003).
21. Carter, D. Visvanathan, N. & Pickles, A. J. The dwarf star content of elliptical and lenticular galaxies. *Astrophys. J.* **311,** 637–650 (1986).
22. Treu, T. *et al.* The initial mass function of early-type galaxies. *Astrophys. J.* **709,** 1195–1202 (2010).
23. Calchi Novati, S. & de Luca, F. Jetzer, Ph. Mancini, L. & Scarpetta, G. Microlensing constraints on the Galactic bulge initial mass function. *Astron. Astrophys.* **480,** 723–733 (2008).
24. van der Wel, A. *et al.* Recent structural evolution of early-type galaxies: size growth from $z = 1$ to $z = 0$. *Astrophys. J.* 688**,** 48–58 (2008).
25. Larson, R. B. Thermal physics, cloud geometry and the stellar initial mass function. *Mon. Not. R. Astron. Soc.* **359,** 211–222 (2005).
26. Bate, M. R. & Bonnell, I. A. The origin of the initial mass function and its dependence on the mean Jeans mass in molecular clouds. *Mon. Not. R. Astron. Soc.* **356,** 1201–1221 (2005).
27. Bonnell, I. A., Clark, P. & Bate, M. R. Gravitational fragmentation and the formation of brown dwarfs in stellar clusters. *Mon. Not. R. Astron. Soc.* **389,** 1556–1562 (2008).
28. Schneider, R. & Omukai, K. Metals, dust and the cosmic microwave background: fragmentation of high-redshift star-forming clouds. *Mon. Not. R. Astron. Soc.* **402,** 429–435 (2010).
29. Banerji, S., Viti, S., Williams, D. A. & Rawlings, J. M. C. Timescales for low-mass star formation in extragalactic environments: implications for the stellar initial mass function. *Astrophys. J.* **692,** 283–289 (2009).
30. Labbé, I. *et al.* Star formation rates and stellar masses of $z = 7$–8 galaxies from IRAC observations of the WFC3/IR Early Release Science and the HUDF fields. *Astrophys. J.* **716,** L103–L108 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Bezanson, J. Brinchmann, R. Larson and R. Zinn for discussions. We thank R. Schiavon for comments that improved the manuscript. This study is based on observations obtained at the W. M. Keck Observatory. We recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community and consider ourselves fortunate to have the opportunity to conduct observations from this mountain.

Author Contributions P.G.v.D. obtained and analysed the data and contributed to the analysis and interpretation. C.C. constructed the stellar population synthesis models and contributed to the analysis and interpretation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.G.v.D. (pieter.vandokkum@yale.edu).

doi:10.1038/nature09580

The mechanism of sodium and substrate release from the binding pocket of vSGLT

Akira Watanabe[1]*, Seungho Choe[2]*, Vincent Chaptal[1], John M. Rosenberg[2,3], Ernest M. Wright[1], Michael Grabe[2,3] & Jeff Abramson[1]

Membrane co-transport proteins that use a five-helix inverted repeat motif have recently emerged as one of the largest structural classes of secondary active transporters[1,2]. However, despite many structural advances there is no clear evidence of how ion and substrate transport are coupled. Here we report a comprehensive study of the sodium/galactose transporter from *Vibrio parahaemolyticus* (vSGLT), consisting of molecular dynamics simulations, biochemical characterization and a new crystal structure of the inward-open conformation at a resolution of 2.7 Å. Our data show that sodium exit causes a reorientation of transmembrane helix 1 that opens an inner gate required for substrate exit, and also triggers minor rigid-body movements in two sets of transmembrane helical bundles. This cascade of events, initiated by sodium release, ensures proper timing of ion and substrate release. Once set in motion, these molecular changes weaken substrate binding to the transporter and allow galactose readily to enter the intracellular space. Additionally, we identify an allosteric pathway between the sodium-binding sites, the unwound portion of transmembrane helix 1 and the substrate-binding site that is essential in the coupling of co-transport.

Secondary active transporters harness the energy stored in electrochemical gradients to drive the accumulation of specific solutes across cell membranes. This task is accomplished by the alternating-access mechanism, in which the substrate-binding site is first exposed to one side of the membrane and, on ion and substrate binding, a conformational change exposes the transported solute to the opposite face, where it is released[3]. Sodium/glucose co-transporters are prototypes of secondary active transporters that drive the accumulation of sugars and other molecules into cells. These transporters have critical roles in human physiology, where mutations in their genes are responsible for severe congenital diseases[4] and are the molecular targets for drugs to treat diabetes and obesity[5].

There has been a recent surge of work on crystal structures[6–11] displaying the five-helix inverted repeat motif. These are referred to as the LeuT superfamily and include genetically diverse proteins that transport a wide range of substrates and differ in the number and type of driving ligand[1,2,12]. A general model for alternating access is being pieced together through comparisons of these diverse structures[1,2,13]. Despite sharing a common set of ten core transmembrane segments, the lack of sequence similarity and the chemical diversity of the transported substrates prevents the complete understanding of the mechanistic basis of transport. This hurdle is being surmounted as multiple structures of the same protein—at different stages in the transport cycle—are solved, providing a comprehensive understanding of substrate binding[7,14–16] and the transition from outward- to inward-facing conformations[17]. However, an atomic-level understanding of sodium-coupled substrate co-transport, necessary to explain the dynamics of alternating access, is still absent.

To investigate the mechanism of sodium–sugar coupling, we carried out molecular dynamics simulations on the galactose-bound inward-occluded conformation of vSGLT[6] embedded in a lipid bilayer[18]. All sodium co-transporters of the LeuT superfamily share a common sodium-binding site termed the Na2 site. During the transition to the inward-facing conformation, transmembrane helix (TM) 8, which forms part of the sodium-binding site, is displaced by ~4 Å, generating a less favourable Na2 site that facilitates Na^+ release[1,17] (Fig. 1b). Na^+ modelled at this site is loosely coordinated by the carbonyl oxygens of Ile 65 (3.3 Å), Ala 361 (3.2 Å) and the side-chain hydroxyl of Ser 365 (3.1 Å). The carbonyl oxygen of Ala 62 (3.6 Å) and the side-chain hydroxyl of Ser 364 (3.6 Å) are also in close proximity (Fig. 1b). Previous molecular dynamics simulations performed on vSGLT[19] and Mhp1[17] indicated that Na^+ quickly leaves the Na2 site. Our simulations indicate that Na^+ exits the Na2 site after 9 ns (Fig. 2a) and interacts with the hydrophilic pore-lining residue Asp 189 on TM5 during exit. The importance of Asp 189 was highlighted in a previous simulation[19] and in biochemical studies on hSGLT1[20]. All three molecular dynamics simulations indicate that Na^+ exits the transporter before substrate exit; however, additional conformational changes are required to release the occluded galactose.

In the inward-occluded structure, galactose is located halfway across the membrane (Figs 1a and 2b), where it is coordinated by extensive side-chain interactions from TM1, TM2, TM6, TM7 and TM10. Subsets of these residues form two hydrophobic gates blocking galactose exit to the intracellular and extracellular spaces. Our molecular dynamics simulations show that as Na^+ exits the Na2 site, galactose undergoes significant fluctuations within the binding pocket. At 52 ns, Tyr 263 adopts a new and stable rotamer conformation that expands the exit pathway (between 52 and 110 ns), permitting the sugar to leave the binding site (Figs 2 and 3). After sugar release (~110 ns), Tyr 263 returns to the original conformation.

To test the hypothesis that Na^+ release stimulates an alternative conformation of Tyr 263, we conducted a 200-ns molecular dynamics simulation in which the sodium was lightly restrained in the Na2 site. Under these conditions, Tyr 263 never adopts the alternative conformation, and thereby prevents galactose exit (Supplementary Fig. 1). This observation suggests that sodium release drives conformational changes that disrupt the galactose-binding site, and further suggests that interactions between the Na2 site and Tyr 263 are central to the transport mechanism.

The spontaneous release of galactose in the absence of applied forces makes possible the accurate determination of the binding free energy profile through the use of umbrella sampling along the exit pathway coupled with weighted histogram analysis[21] (Fig. 3, inset). After Na^+ release, galactose is weakly bound to vSGLT with a minimal energy barrier of ~2 kcal mol^{-1}, resulting from the interaction of the sugar with residues Asn 64, Ser 66, Glu 68 and Gln 69 on TM1. Asn 64 is of particular interest because it is located in the unwound segment of TM1 and has hydrogen bonds with the inner gate residue Tyr 263 and the O2 hydroxyl of galactose linking the Na2 site with the galactose site. Thus, the interactions of Asn 64 with Tyr 263 and galactose may be critical to the transport mechanism[6].

To test the importance of these interactions, we performed molecular dynamics simulations and sodium-dependent transport assays on

[1]Department of Physiology, University of California, Los Angeles, Los Angeles, California 90095-1759, USA. [2]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA. [3]Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA.
*These authors contributed equally to this work.

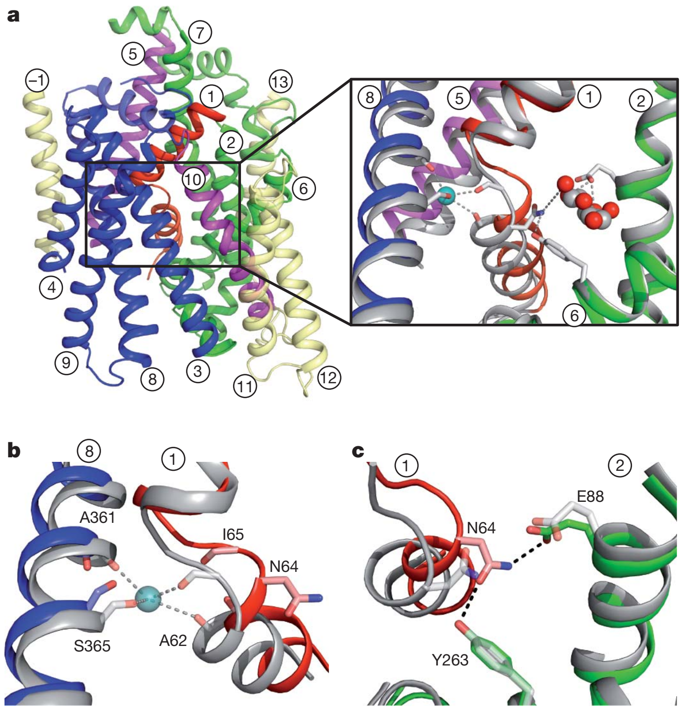


Figure 1 | Structures and overlay of the inward-open and inward-occluded conformations. **a**, The core domain of the inward-open conformation (TM1–TM10) is coloured by specific helix bundles involved in the transition from the inward-occluded to the inward-open conformation. The 'hash motif' formed from TM3, TM4, TM8 and TM9 is blue; the 'sugar bundle' formed from TM2, TM6 and TM7 is green; TM1 is red; and TM5 and TM10 are magenta. The periphery helices (TM−1, TM11, TM12 and TM13) are yellow. Atoms are displayed in ball-and-stick form with oxygen coloured red and nitrogen coloured blue. Inset, an overlay of the inward-open (colour) and inward-occluded (grey) conformations illustrating the coordination at the Na2 and galactose-binding sites. **b**, **c**, Overlay of the inward-open and inward-occluded conformations with the same colouring as in **a**. Conformational changes in the inward-open structure reveals a ~13° kink in the unwound segment of TM1 that prevents sodium coordination at the Na2 site (**b**). In the absence of galactose, the galactose-binding residue Asn 64 hydrogen-bonds to Glu 88 and Tyr 263, maintaining an open pathway from the intracellular space to the substrate-binding site (**c**).

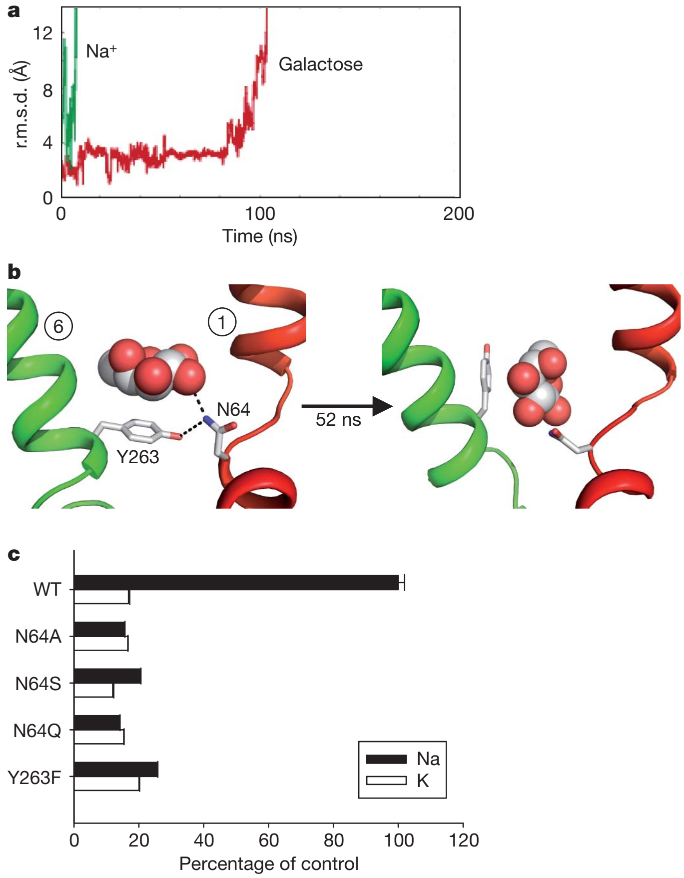


Figure 2 | Mechanism of galactose release. **a**, Sodium and galactose exit vSGLT. The root mean squared deviation (r.m.s.d.) of Na^+ (green) rapidly increases at 9 ns, indicating exit from the Na2 site. This is followed by the release of galactose (red) at 110 ns. **b**, Tyr 263 adopts two rotamers. On the left, Tyr 263 is shown in the conformation observed in the inward-occluded structure[6], in which it blocks substrate exit through a hydrogen bond with Asn 64 on TM1. At 52 ns (shown on the right), Tyr 263 adopts a rotamer conformation that expands the exit pathway. **c**, D-galactose uptake by wild-type and vSGLT mutants in proteoliposomes. Results are expressed as percentage uptake in either 100 mM NaCl or KCl, and show that the mutants Asn 64 Ala, Asn 64 Ser, Asn 64 Gln and Tyr 263 Phe severely impair sodium-dependent transport. Error bars, s.e.m. WT, wild type.

transporters with mutations at positions 64 and 263. Simulations of the Asn 64 Ala mutant show a momentary sodium departure from the Na2 site at 5 ns, but the ion rapidly returns and remains for the remainder of the simulation. The failure of Na^+ to unbind prevents conformational changes in the unwound segment of TM1, and Tyr 263 remains in the blocked orientation (Supplementary Fig. 2a). In agreement with the simulation, sodium-dependent transport assays on the Asn 64 Ala mutant show no activity (Fig. 2c).

To explore the role of Asn 64 further, we tested Asn 64 Gln and Asn 64 Ser, which, in principal, are both capable of maintaining the native hydrogen bonds to Tyr 263 and galactose. Models of Asn 64 Gln prevented simulation as the result of substantial steric clashes, which correlated well with a lack of transport (Fig. 2c). The model of the Asn 64 Ser mutation could form a hydrogen bond with Tyr 263 (3.3 Å) but not with the O2 hydroxyl of galactose (4.3 Å). In the simulation, Asn 64 Ser releases Na^+ from the Na2 site at 25 ns, and Tyr 263 transiently adopts the alternative rotamer conformation before returning to its original position, preventing galactose exit (Supplementary Fig. 2c). Similarly, simulation of the Tyr 263 Phe mutant shows that Na^+ unbinds at 10 ns, but unlike tyrosine, phenylalanine never adopts a conformation compatible with galactose exit (Supplementary Fig. 2b). Longer simulations may reveal galactose release in these mutants, because they both show modest transport activity (~10% of wild type); however, the transport assays and simulation data demonstrate that robust transport requires precise orientation of Asn 64 to stabilize galactose and the gating residue Tyr 263.

Although the molecular dynamics simulations and biochemical studies demonstrate a physical link between the Na2 site and the substrate, global details regarding the inward-open conformation (devoid of both ligands) remain elusive. To address this issue, we determined the structure of vSGLT in the inward-open conformation. Crystals, in the absence of ligands, for both the wild-type protein and the inactive Lys 294 Ala mutant[6] were obtained. Both crystals had the same overall configuration, but the mutant crystals diffracted to a higher resolution (2.7 Å; see Methods).

As in the original structure[6], the inward-open conformation is composed of 14 transmembrane helices, ten of which comprise the core domain. TM1–TM5 and TM6–TM10 are related by an approximate two-fold symmetry axis through the centre of the membrane plane. The inward-occluded and the inward-open structures have a similar overall fold with a r.m.s.d. of 1.2 Å. However, there are distinct structural differences between the two conformations, presumably owing to changes resulting from the release of ligands (Fig. 4). With the exception of TM1, superimpositions of individual helices reveal that the occluded-to-open transition occurs by rigid-body movements of subdomains (Supplementary Figs 3 and 4). Consistent with the recent

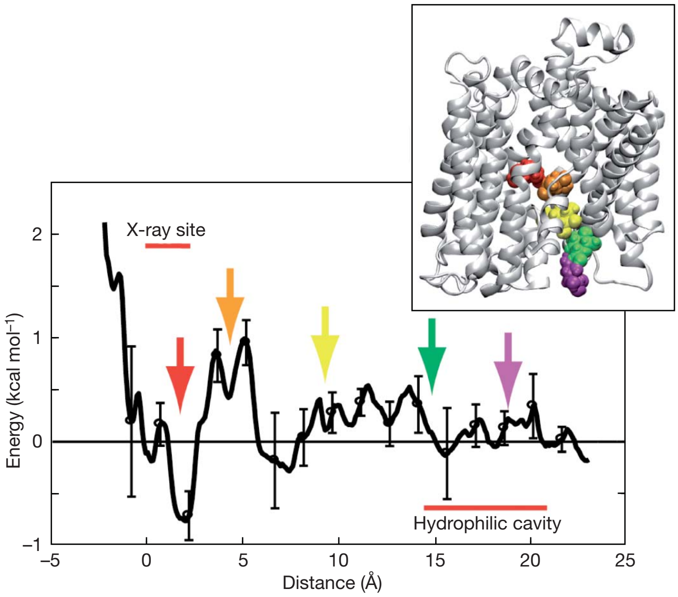


Figure 3 | The potential of mean force for galactose unbinding. Energy of galactose binding to vSGLT in the absence of Na^+. Umbrella sampling along the natural, equilibrium pathway shown (inset) was used to determine the binding free energy. The distance along the pathway from the binding site in the X-ray structure is shown along the *x* axis. The coloured arrows correspond to the galactose positions shown in the inset. The largest barrier is ~2 kcal mol^{-1}, at 5 Å, which corresponds to galactose interaction with residues in the kink region of TM1. Error bars were determined by splitting the production data into four equal sets, computing the energy profile for each set, and then applying a global shift to each curve before calculating the standard deviation at the 16 positions marked with points.

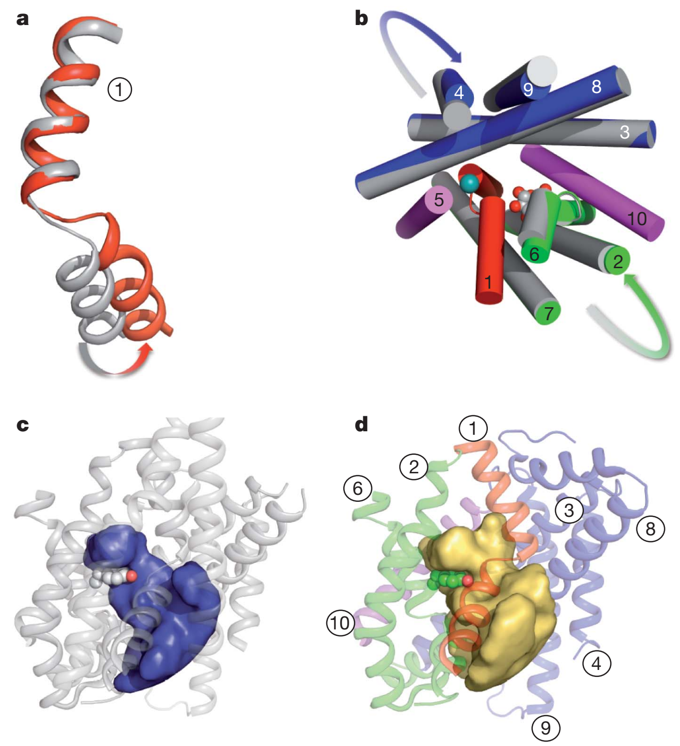


Figure 4 | Conformational changes in the transition from the inward-occluded to the inward-open structure. a, TM1 superimposed between the inward-open (red) and inward-occluded (grey) structures, showing a ~13° kink in TM1. **b**, Overlay of the inward-open (coloured as in Fig. 1) and inward-occluded (grey) conformations. Rigid-body rotations of the hash motif and sugar bundle by 3° in opposite directions expose the substrate-binding site to the intracellular environment. **c**, Accessibility cavity of the inward-occluded conformation is coloured blue. **d**, Accessibility cavity of the inward-open conformation is coloured gold. The conformational changes from TM1, hash motif and sugar bundle cause an increase of ~1,400 Å^3 in the accessible volume of the inward-open conformation, aiding galactose release.

assignment for Mhp1[17], the hash motifs, formed from TM3 and TM4 and their inverted repeat equivalents, TM8 and TM9, align with a r.m.s.d. of 0.9 Å. TM2, TM6 and TM7 form a domain termed the sugar bundle for the extensive side-chain interactions with galactose, and these regions superimpose with a r.m.s.d. of 0.5 Å. This new inward-open structure of vSGLT is more similar to the recent inward-facing conformation structure of Mhp1 than is the previous structure of vSGLT. Details of this structural analysis are in Methods.

The transition from the inward-occluded to the inward-open structure is presumably triggered by sodium release from the Na2 site and the alteration in the hydrogen-bonding network surrounding the unwound segment of TM1. In particular, the intracellular half of TM1 flexes ~13°, modifying the coordination of Asn 64 (Figs 1b and 4a). In the absence of both galactose and Na^+, Asn 64 coordinates Tyr 263 and Glu 88. Glu 88 was previously hydrogen-bonded to the O2 and O3 hydroxyls of galactose. This new conformation of TM1 is further stabilized by hydrogen bonds between the Na2-site residue Ser 365 and Glu 68 on the unwound segment of TM1 (Supplementary Fig. 5). When viewed from the intracellular side, each domain moves ~3° in opposite directions, thereby increasing the volume of the accessibility cavity by ~1,400 Å^3 (Fig. 4). This 6° relative rotation probably disrupts protein–substrate coordination and permits water to enter the site. This interpretation is supported by our simulations, in which an increase in the number of water molecules in the substrate-binding site is observed after sodium release (Supplementary Fig. 6). Water effectively competes with the protein for hydrogen bonds, loosening galactose in the pocket and ultimately assisting in its release (Supplementary Fig. 6 and Supplementary Movie 1).

We propose the following mechanism for sodium and galactose exit from vSGLT. The transition from the outward- to the inward-occluded conformation weakens the Na2 sodium-binding site, causing it to become metastable and release the ion on a short timescale. Upon exit, the carbonyl oxygens of the ion-coordinating residues Ile 65 and Ala 62 undergo a conformational change in the unwound segment of TM1, producing a kink of ~13° (Figs 1 and 4a). Our simulation shows that movement of TM1 disrupts the hydrogen bond between Asn 64 and Tyr 263, allowing the side chain of Tyr 263 to adopt a new conformation that opens a pathway to the intracellular space (Fig. 2b). Additional rigid-body movements widen the intracellular cavity, allowing water penetration and further disrupting the substrate-binding site to enhance exit and prevent rebinding (Fig. 4).

It is likely that the reaction scheme described here for vSGLT is broadly used by all sodium-dependent members of the LeuT superfamily, because the Na2 site, the hydrophobic gates and the unwound segments on TM1 and TM6 are all conserved[1,11,22,23]. For proteins with a single sodium-binding site, the Na2 site directly interacts with the substrate through polar residues—Asn 64 in vSGLT and Gln 42 in Mhp1—located on the unwound segment of TM1. For proteins that harbour two sodium-binding sites, such as LeuT and, putatively, BetP, the additional site (the Na1 site) is positioned on the opposite side of the unwound helix from the Na2 site. Interactions between the Na1 and Na2 sites are mediated by the unwound segment of TM1, and the sodium at the Na1 site is directly coordinated to the substrate[8,12]. Regardless of whether the protein has one or two sodium-binding sites, it is the conserved Na2 site, positioned most distal from the core of the protein, that regulates sodium and substrate release. This primary structural feature coupling sodium and substrate co-transport has fundamental implications for our understanding of membrane protein biology and for developing strategies to manipulate the alternating-access mechanism therapeutically.

METHODS SUMMARY

Molecular dynamics simulations. The vSGLT monomer (Protein Data Bank ID, 3DH4) was embedded and solvated in a 1-palmitoyl-2-oleoyl phosphatidylcholine membrane bilayer using the OPM[24] and CHARMM-GUI[18] software packages. Simulations were carried out using NAMD[25] with the CHARMM27 parameter set in a 150 mM NaCl bath. See Methods for more details.

Protein expression and purification. Plasmids carrying wild-type or mutant transporters were transformed and overexpressed in the TOP10 *Escherichia coli* cell line. Cell membranes were isolated, solubilized (2% w/v decyl-β-D-maltopyranoside) and tandem-purified using a Ni-NTA Superflow column (affinity chromatography) and a Superdex 200 column (size-exclusion chromatography). See Methods for more details.

Transport assays. We generated proteoliposomes by reconstituting purified vSGLT protein with sonicated lipid at a protein/lipid ratio of 1:200. We measured transport activity by monitoring the uptake of D-galactose, with ^{14}C-D-galactose tracer, into proteoliposomes in the presence or absence of a 100 mM Na^+ gradient (K^+ replacing Na^+). See Methods for more details.

Crystallization and data collection. We concentrated purified wild-type and Lys 294 Ala protein to $\sim$13 mg ml^{-1} and grew crystals by the hanging-drop vapour diffusion method using the Mosquito nanolitre-dispensing robot. Data collected at the Advanced Light Source, Berkeley (beamline 5.0.2), were integrated and scaled, and phases were calculated by molecular replacement. The model was built and refined to an R_{work}/R_{free} value of 25.1/27.4. See Methods for more details.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 May; accepted 12 October 2010.
Published online 5 December 2010.

1. Abramson, J. & Wright, E. M. Structure and function of Na^+-symporters with inverted repeats. *Curr. Opin. Struct. Biol.* **19,** 425–432 (2009).
2. Krishnamurthy, H., Piscitelli, C. L. & Gouaux, E. Unlocking the molecular secrets of sodium-coupled transporters. *Nature* **459,** 347–355 (2009).
3. Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211,** 969–970 (1966).
4. Wright, E. M., Hirayama, B. A. & Loo, D. F. Active sugar transport in health and disease. *J. Intern. Med.* **261,** 32–43 (2007).
5. Isaji, M. Sodium-glucose cotransporter inhibitors for diabetes. *Curr. Opin. Investig. Drugs* **8,** 285–292 (2007).
6. Faham, S. *et al.* The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+/sugar symport. *Science* **321,** 810–814 (2008).
7. Weyand, S. *et al.* Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science* **322,** 709–713 (2008).
8. Ressl, S., Terwisscha van Scheltinga, A. C., Vonrhein, C., Ott, V. & Ziegler, C. Molecular basis of transport and regulation in the Na^+/betaine symporter BetP. *Nature* **458,** 47–52 (2009).
9. Fang, Y. *et al.* Structure of a prokaryotic virtual proton pump at 3.2 Å resolution. *Nature* **460,** 1040–1043 (2009).
10. Gao, X. *et al.* Structure and mechanism of an amino acid antiporter. *Science* **324,** 1565–1568 (2009).
11. Shaffer, P. L., Goehring, A., Shankaranarayanan, A. & Gouaux, E. Structure and mechanism of a Na^+-independent amino acid transporter. *Science* **325,** 1010–1014 (2009).
12. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters. *Nature* **437,** 215–223 (2005).
13. Forrest, L. R. & Rudnick, G. The rocking bundle: a mechanism for ion-coupled solute flux by symmetrical transporters. *Physiology (Bethesda)* **24,** 377–386 (2009).
14. Zhou, Z. *et al.* LeuT-desipramine structure reveals how antidepressants block neurotransmitter reuptake. *Science* **317,** 1390–1393 (2007).
15. Gao, X. *et al.* Mechanism of substrate recognition and transport by an amino acid antiporter. *Nature* **463,** 828–832 (2010).
16. Singh, S. K., Piscitelli, C. L., Yamashita, A. & Gouaux, E. A competitive inhibitor traps LeuT in an open-to-out conformation. *Science* **322,** 1655–1661 (2008).
17. Shimamura, T. *et al.* Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1. *Science* **328,** 470–473 (2010).
18. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29,** 1859–1865 (2008).
19. Li, J. & Tajkhorshid, E. Ion-releasing state of a secondary membrane transporter. *Biophys. J.* **97,** L29–L31 (2009).
20. Quick, M., Loo, D. D. & Wright, E. M. Neutralization of a conserved amino acid residue in the human Na^+/glucose transporter (hSGLT1) generates a glucose-gated H^+ channel. *J. Biol. Chem.* **276,** 1728–1734 (2001).
21. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A. & Rosenberg, J. M. The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J. Comput. Chem.* **13,** 1011–1021 (1992).
22. Yernool, D., Boudker, O., Jin, Y. & Gouaux, E. Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*. *Nature* **431,** 811–818 (2004).
23. Hunte, C. *et al.* Structure of a Na^+/H^+ antiporter and insights into mechanism of action and regulation by pH. *Nature* **435,** 1197–1202 (2005).
24. Lomize, A. L., Pogozheva, I. D., Lomize, M. A. & Mosberg, H. I. Positioning of proteins in membranes: a computational approach. *Protein Sci.* **15,** 1318–1333 (2006).
25. Kale, L. *et al.* NAMD2: Greater scalability for parallel molecular dynamics. *J. Comput. Phys.* **151,** 283–312 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Vondriska and K. Philipson as well as members of the Abramson, Wright and Grabe labs for useful discussions and for reading the manuscript. We would also like to thank S. Faham for contributions at the early stages of this work, S. Iwata for advance release of the Mhp1 coordinates (Protein Data Bank ID, 2X79), and R. Roskies for assistance with the computations. Simulations were carried out through a TeraGrid grant at the Pittsburgh Supercomputing Center and the Texas Advanced Computing Center. This work was supported by NIH grants GM078844 (J.A.), RGY0069 (J.A.) and DK19567 (E.M.W.), and a grant from the Human Frontier Science Program (J.A.). M.G. is an Alfred P. Sloan Research Fellow.

Author Contributions Experiments were carried out and diffraction data collected by A.W., V.C. and J.A. Simulations were carried out by S.C. Data were analysed and the manuscript was prepared by all authors.

Author Information Coordinates and structure factors of the inward-open vSGLT structure have been deposited in the Protein Data Bank under accession number 2XQ2. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.A. (jabramson@mednet.ucla.edu) or M.G. (mdgrabe@pitt.edu).

METHODS

Molecular dynamics simulations. Initially, TM−1 was removed and six missing residues in the TM4–TM5 loop were added with the loop modelling routine in Modeller[26]. Residues 53–547 were then embedded in a membrane and solvated in a hexagonal box approximately $96 \times 96 \times 84$ Å^3 in volume for a total of 63,000 atoms. Electroneutrality was enforced with the addition of 150 mM NaCl.

Simulations were carried out with the CMAP corrected[27] CHARMM27 parameter set and the TIP3P water model. VMD[28] and MATLAB were used for visualization and analysis. The system was minimized using conjugate gradient minimization and heated to 310 K using Langevin dynamics with a 10-ps^{-1} damping coefficient. An initial 300-ps equilibration using the NVT ensemble was carried out in which water, galactose, Na^+ and heavy backbone and side-chain atoms were constrained in a harmonic potential with a force constant of $k = 10.0$ kcal mol^{-1} Å^{-2}. We then switched to an NPT ensemble, and the restraints on the water molecules and heavy side-chain atoms were gradually removed in five steps over 1.5 ns. All remaining restraints were removed in six steps over the next 1.8 ns. Finally, 10 ns of restraint-free simulation was run. All production runs start from this equilibrated system. A Langevin piston with a 200-fs period and 100-fs decay was used to set the pressure to 1 atm. Hydrogen bond lengths were constrained with SHAKE[29], and a 2-fs time step was used. A 10 Å van der Waals cut-off was used along with the particle mesh Ewald method for the electrostatics.

Simulations with the restrained ion were carried out by holding the Na^+ in a weak spherical harmonic potential using the distance from the Na^+ to the centre of mass (COM) of the of the five coordinating residues. The equilibrium distance, based on the inward-occluded structure, was 1.39 Å, and a force constant of $k = 0.5$ kcal mol^{-1} Å^{-2} produced minimal distortions in the protein. A 10-ns equilibration was run before a 200-ns production run. The r.m.s.d. of the unrestrained simulation was 2.5 Å for the entire protein and 3.0 Å for the restrained-Na^+ simulation.

Potential of mean force calculation. The potential of mean force (PMF) was calculated using umbrella sampling with WHAM[21]. We extracted 69 snapshots along the pathway and held each configuration in a harmonic potential ($k = 7.0$ kcal mol^{-1} Å^{-2}) with a resting length equal to the z component of the distance between the galactose COM and the binding-site COM defined by the binding-site residues. Two nanosecond trajectories were run for each umbrella, and the last 1,800 ps were used for calculating the PMF. Splitting the trajectories into two equal parts (200–1,000 ps and 1,200–2,000 ps) and computing separate PMFs revealed that the total PMF is well converged.

Protein purification. vSGLT proteins were cloned, expressed and purified as previously described[6,30–32]. Briefly, the plasmids were transformed into the TOP10 cell line expressed to OD_{600} 1.8 and induced with 0.66 mM L-arabinose for 4 h at 29 °C. Cell membranes were isolated, solubilized with 2% decyl-β-D-maltopyranoside and affinity-purified on a Ni-NTA column. The sample was further purified by size-exclusion chromatography (Superdex 200) and washed with crystal buffer (20 mM Tris (pH 7.5), 25 mM NaCl, 0.174% decyl-β-D-maltopyranoside) in a 50-kDa Amicon filter unit.

Transport assays. Mutants were created with the QuikChange method and purified as above. vSGLT protein was reconstituted in 150 mM KCl, 10 mM Tris/Hepes (pH 8.0), 1 mM DTT, 1 mM Na_2EDTA, 1 mM $CaCl_2$, 1 mM $MgCl_2$ and 0.5% decyl-β-D-maltopyranoside, with 1.2 mg ml^{-1} sonicated lipid (90 mg asolectin soy lecithin, 10 mg cholesterol) at a protein/lipid ratio of 1:200. Addition of 5 mg ml^{-1} SM-2 Bio-Beads initiated the reconstitution and the mixture was incubated overnight at 4 °C. The proteoliposomes were collected and washed twice by centrifugation. Pelleted proteoliposomes were resuspended and underwent three freeze–thaw cycles in liquid nitrogen.

Uptake of D-galactose (88 μM) with ^{14}C-D-galactose tracer into proteoliposomes was measured for 18 min at 22 °C in the presence or absence of 100 mM Na^+ (K^+ replacing Na^+) as described previously[6,30]. Proteoliposomes were collected by filtration through 0.45-μm Millipore filters and the uptake was quantified by scintillation counting. Results are expressed as the mean ± s.e.m. of three determinations and three trials.

Crystallization. Protein was concentrated to ~13 mg ml^{-1} before plating. Optimization by additive screening gave the best diffracting crystals with a reservoir solution containing 0.1 M MES (pH 6.5), 4% MPD and 9–13% PEG400, and tridecyl-β-D-maltopyranoside to a final concentration of 0.0017% as an additive. Before freezing, crystals were cryoprotected using a solution containing 30% PEG400 and 0.174% decyl-β-D-maltopyranoside.

Data processing, phasing and refinement. Data was collected at 1.0 Å on cryo-cooled crystal (100 K) at the Advanced Light Source (beamline 5.0.2). Five data sets were integrated using HKL2000[33] and merged and subjected to B-factor-sharpening using an anisotropy correction server[34] (resolution cut-offs: $a = 3.1$ Å, $b = 2.7$ Å and $c = 2.8$ Å). Phases were calculated by molecular replacement (PHASER[35]) using the original vSGLT structure as a search model. The model was built in COOT[36] and refined using PHENIX[37] and BUSTER[38] using non-crystallographic symmetry (NCS) and TLS refinement restraints. There are two molecules per asymmetric unit with the A molecule displaying sharper electron density and lower B factors (88.5 Å^2) than the B molecule (131.3 Å^2). The model was built and refined to an R_{work}/R_{free} value of 25.1/27.4. The Ramachandran statistics shown areas follows: 95.5% of the residues lie in the preferred region, 4.3% lie in the allowed region and 0.2% are outliers.

The $2F_o$–F_c maps contained three elongated features having a maximal peak height of 3σ. These attributes were interpreted and assigned as PEG molecules. Two are located at the periphery, whereas the third is near the Na2 site as observed in the Mhp1 structure[17] and is proposed to stabilize the inward-facing conformation.

The Lys 294 Ala protein crystals diffract to higher resolution than the wild-type crystals. Data from four wild-type crystals were collected and merged to achieve a 3.7 Å resolution data set. Difference Fourier maps were calculated against the final Lys 294 Ala mutant model and no significant peaks were observed. The Lys 294 Ala model was further refined using PHENIX to yield an R_{work}/R_{free} value of 30.7/34.8.

We note that although refinement was carried out with data subject to anisotropic correction, as described above, the deposited data has not been treated. Figures were created from the A-chain protomer using PYMOL[39].

Structural comparison of vSGLT with Mhp1. Superpositions of the inward-occluded (Protein Data Bank ID, 3DH4) and inward-open conformations of vSGLT with the inward-facing conformation of Mhp1 (Protein Data Bank ID, 2X79) reveal they all share a similar global fold. The largest differences are centred near the substrate- and ion-binding sites. The Na2-site helices (TM1, TM5 and TM8) of the inward-open conformation of vSGLT have a closer fit to Mhp1 (r.m.s.d., 2.2 Å) than the inward-occluded conformation (r.m.s.d., 2.6 Å); thus, the inward-open vSGLT structure more closely resembles the structure of Mhp1.

26. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234,** 779–815 (1993).
27. MacKerell, A. D. Jr, Feig, M. & Brooks, C. L. III. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **126,** 698–699 (2004).
28. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).
29. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **23,** 327–341 (1977).
30. Turk, E. *et al.* Molecular characterization of Vibrio parahaemolyticus vSGLT: a model for sodium-coupled sugar cotransporters. *J. Biol. Chem.* **275,** 25711–25716 (2000).
31. Turk, E., Gasymov, O. K., Lanza, S., Horwitz, J. & Wright, E. M. A reinvestigation of the secondary structure of functionally active vSGLT, the Vibrio sodium/galactose cotransporter. *Biochemistry* **45,** 1470–1479 (2006).
32. Veenstra, M., Lanza, S., Hirayama, B. A., Turk, E. & Wright, E. M. Local conformational changes in the Vibrio Na^+/galactose cotransporter. *Biochemistry* **43,** 3620–3627 (2004).
33. Otwinowski, Z. & Minor, W. in *Macromolecular Crystallography* (eds Abelson, J. N., Simon, M. I., Carter, C. W. Jr & Sweet, R. M.) 307–326 (Methods in Enzymology 276, Academic, 1997).
34. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc. Natl Acad. Sci. USA* **103,** 8060–8065 (2006).
35. Qian, B. *et al.* High-resolution structure prediction and the crystallographic phase problem. *Nature* **450,** 259–264 (2007).
36. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
37. Adams, P. D. *et al.* Recent developments in the PHENIX software for automated crystallographic structure determination. *J. Synchrotron Radiat.* **11,** 53–55 (2004).
38. Joosten, R. P., Womack, T., Vriend, G. & Bricogne, G. Re-refinement from deposited X-ray data can deliver improved models for most PDB entries. *Acta Crystallogr. D* **65,** 176–185 (2009).
39. DeLano, W. L. *PyMOL Molecular Viewer* ⟨http://www.pymol.org⟩ (2002).

doi:10.1038/nature09587

Subtypes of medulloblastoma have distinct developmental origins

Paul Gibson[1], Yiai Tong[1], Giles Robinson[1,2], Margaret C. Thompson[9], D. Spencer Currle[1], Christopher Eden[1], Tanya A. Kranenburg[1], Twala Hogg[1], Helen Poppleton[1], Julie Martin[1], David Finkelstein[3], Stanley Pounds[4], Aaron Weiss[10], Zoltan Patay[5], Matthew Scoggins[5], Robert Ogg[5], Yanxin Pei[11], Zeng-Jie Yang[11], Sonja Brun[11], Youngsoo Lee[6], Frederique Zindy[6], Janet C. Lindsey[12], Makoto M. Taketo[13], Frederick A. Boop[7], Robert A. Sanford[7], Amar Gajjar[2], Steven C. Clifford[12], Martine F. Roussel[6], Peter J. McKinnon[6], David H. Gutmann[14], David W. Ellison[8], Robert Wechsler-Reya[11] & Richard J. Gilbertson[1,2]

Medulloblastoma encompasses a collection of clinically and molecularly diverse tumour subtypes that together comprise the most common malignant childhood brain tumour[1–4]. These tumours are thought to arise within the cerebellum, with approximately 25% originating from granule neuron precursor cells (GNPCs) after aberrant activation of the Sonic Hedgehog pathway (hereafter, SHH subtype)[3–8]. The pathological processes that drive heterogeneity among the other medulloblastoma subtypes are not known, hindering the development of much needed new therapies. Here we provide evidence that a discrete subtype of medulloblastoma that contains activating mutations in the WNT pathway effector *CTNNB1* (hereafter, WNT subtype)[1,3,4] arises outside the cerebellum from cells of the dorsal brainstem. We found that genes marking human WNT-subtype medulloblastomas are more frequently expressed in the lower rhombic lip (LRL) and embryonic dorsal brainstem than in the upper rhombic lip (URL) and developing cerebellum. Magnetic resonance imaging (MRI) and intra-operative reports showed that human WNT-subtype tumours infiltrate the dorsal brainstem, whereas SHH-subtype tumours are located within the cerebellar hemispheres. Activating mutations in *Ctnnb1* had little impact on progenitor cell populations in the cerebellum, but caused the abnormal accumulation of cells on the embryonic dorsal brainstem which included aberrantly proliferating Zic1^{+} precursor cells. These lesions persisted in all mutant adult mice; moreover, in 15% of cases in which *Tp53* was concurrently deleted, they progressed to form medulloblastomas that recapitulated the anatomy and gene expression profiles of human WNT-subtype medulloblastoma. We provide the first evidence, to our knowledge, that subtypes of medulloblastoma have distinct cellular origins. Our data provide an explanation for the marked molecular and clinical differences between SHH- and WNT-subtype medulloblastomas and have profound implications for future research and treatment of this important childhood cancer.

SHH-subtype medulloblastoma is characterized by aberrant SHH signalling that is often driven by inactivating mutations in *PTCH1*[3,4]. These medulloblastomas tend to arise in very young children, display a 'large cell-anaplastic' or 'desmoplastic' histology and have a relatively poor prognosis[2–4]. WNT-subtype medulloblastomas are strikingly different. Arising in much older children, these highly curable tumours have 'classic' morphology and activating mutations in *CTNNB1*[1–4]. Mouse models have shown that SHH-subtype medulloblastomas arise from committed GNPCs of the cerebellum[7,8] and enabled the development of new therapies that suppress the oncogenic SHH-signal[9,10]. It has been suggested that the other medulloblastoma subtypes might have a different cellular origin[5,11,12], but little is known about their biology and there are no mouse models of these tumours.

Recently, we showed that subtypes of the brain tumour ependymoma arise from discrete populations of neural progenitor cells with which they share similar gene expression profiles[13]. Therefore, to determine if medulloblastoma subtypes also arise from discrete cell populations, we first used four online gene expression databases to chart the regional expression of 110 genes that mark human SHH- or WNT-subtype medulloblastomas[3]. Twenty-four WNT-subtype and 25 SHH-subtype medulloblastoma signature genes are contained within 'Brain Explorer 2', which generates three-dimensional gene expression maps across the mouse brain (www.brain-map.org, Supplementary Methods and Supplementary Data set 1). As expected[14], these data revealed the URL at embryonic day (E) 11.5 and the cerebellum at E15.5 to be the most common sites of SHH-subtype signature gene expression (Fig. 1a, b and Supplementary Data Set 1). In contrast, WNT-subtype medulloblastoma signature genes were predominantly expressed within the LRL at E11.5 (rhombomeres (r) 2–r8) and the dorsal brainstem at E15.5. Expression of an additional 61 medulloblastoma signature genes, reported by three other online databases, confirmed this differential pattern (Supplementary Fig. 1 and Supplementary Table 1). These data suggest that SHH- and WNT-subtype medulloblastomas arise from distinct regions of the hindbrain and identify the dorsal brainstem as a potential source of WNT-subtype tumours.

If SHH- and WNT-subtype medulloblastomas have different origins, we reasoned that these tumours should demonstrate anatomical differences at diagnosis. Remarkably, all validated WNT-subtype medulloblastomas examined ($n = 6/6$, Supplementary Fig. 2) were located within the IV ventricle and infiltrated the dorsal surface of the brainstem, whereas all SHH-subtype tumours ($n = 6/6$) were distributed away from the brainstem within the cerebellar hemispheres (Fig. 1c, d and Supplementary Fig. 3, exact Mann–Whitney $P < 0.005$). Five of the six WNT-subtype, but no SHH-subtype, tumours were adherent to the dorsal brainstem at surgery (Fisher's exact test, $P < 0.005$). Thus WNT-subtype medulloblastomas are anatomically distinct from SHH tumours and are intimately related to the IV ventricle and dorsal brainstem.

We noted various cell types surrounding the IV ventricle that could give rise to WNT-subtype medulloblastomas, including dorsal

[1]Department of Developmental Neurobiology, St Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA. [2]Department of Oncology, St Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA. [3]Hartwell Center for Bioinformatics and Biotechnology, St Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA. [4]Department of Biostatistics, St Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA. [5]Department of Radiological Sciences, St Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA. [6]Department of Genetics and Tumor Cell Biology, St Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA. [7]Department of Surgery, St Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA. [8]Department of Pathology, St Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA. [9]Department of Pediatric Hematology/Oncology, The Cleveland Clinic, 9500 Euclid Avenue/S20, Cleveland, Ohio 44195, USA. [10]Robert Wood Johnson Medical School, 195 Little Albany Street, New Brunswick, New Jersey 08903, USA. [11]Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina 27710, USA. [12]Northern Institute for Cancer Research, Newcastle University, Newcastle upon Tyne NE2 4HH, UK. [13]Graduate School of Medicine, Kyoto University, Yoshida-Konoé-cho, Sakyo, Kyoto 606-8501, Japan. [14]Department of Neurology, Washington University School of Medicine, St Louis, Missouri 63110, USA.

brainstem progenitors of cochlear, mossy-fibre and climbing-fibre neurons (Fig. 1a, b and Supplementary Fig. 4)[15]. But it remained possible that cerebellar ventricular-zone radial glia[12,16] or GNPCs generate WNT-subtype medulloblastomas. To identify hindbrain cells that are susceptible to transformation by *Ctnnb1*, we generated mice carrying a cre-dependent mutant allele of *Ctnnb1* ($Ctnnb1^{lox(ex3)}$)[17] and the *Blbp-Cre* transgene[18]. *Blbp-Cre* induces efficient recombination in progenitor cell populations across the hindbrain including the cerebellar ventricular zone, GNPCs of the external germinal layer (EGL) and Olig3^{+} progenitor cells in the LRL[19] (Supplementary Fig. 5). We also generated $Blbp\text{-}Cre^{+/-}$;$Ctnnb1^{+/lox(ex3)}$ (hereafter, *Ctnnb1*-mutant) mice that were homozygous for a cre-dependent mutant allele of *Tp53* ($Tp53^{flx}$)[20] because loss of this tumour suppressor accelerates medulloblastoma formation in $Ptch1^{+/-}$ mice[21]. As expected, *Ctnnb1*-mutant embryos expressed mutant nuclear-Ctnnb1 in all hindbrain germinal zones, regardless of *Tp53* status (Supplementary Figs 5k and 6). Surprisingly, mutation of *Ctnnb1* did not affect significantly the proliferation or apoptosis of ventricular-zone cells or GNPCs in the cerebellum (Fig. 2a and Supplementary Fig. 7).

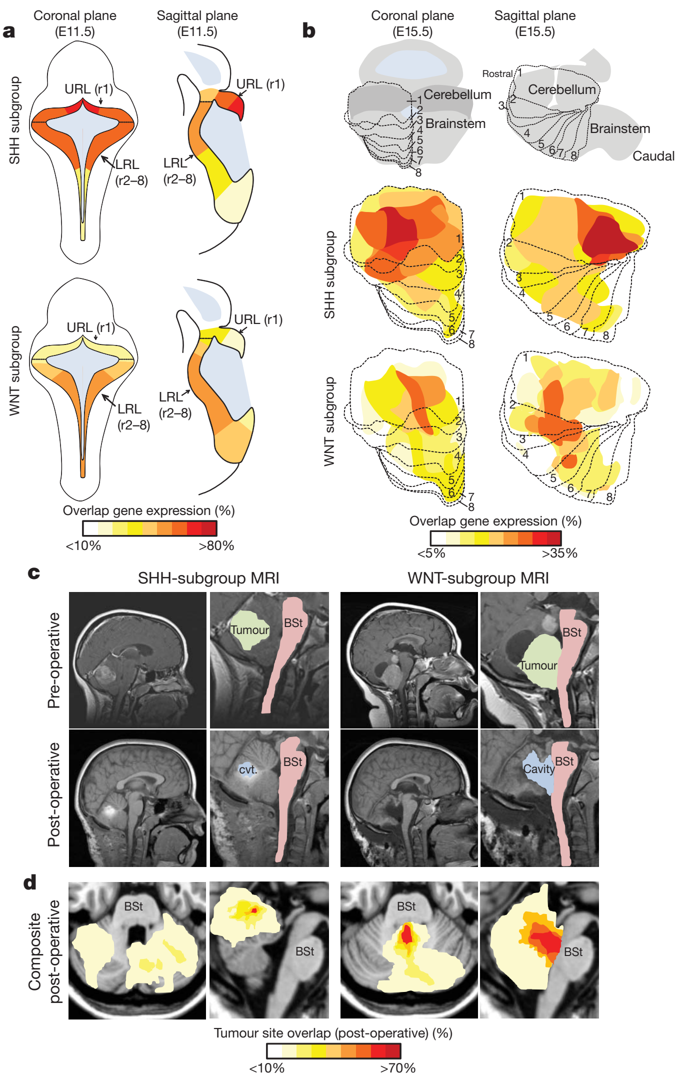


Figure 1 | WNT and SHH subtypes of medulloblastoma are anatomically distinct. **a**, **b**, Expression distribution in (**a**) E11.5 and (**b**) E15.5 mouse hindbrain of orthologues that distinguish human WNT- and SHH-subtype medulloblastoma (Supplementary Data Set 1). Cartoons in **b** denote the position of rhombomeres relative to the cerebellum and brainstem. **c**, Top, pre-operative, and bottom, post-operative, MRI scans of exemplary SHH- and WNT-subtype medulloblastomas. Right panels show close-up views of left panels. Brainstem (BSt), post-operative tumour cavity (cvt). **d**, Frequency and site of post-operative surgical cavities of SHH- ($n = 6$) and WNT- ($n = 6$) subtype medulloblastomas. Axial (left) and sagittal (right) views are shown.

Because GNPCs generate SHH-subtype medulloblastomas[7,8], we sought additional evidence that these cells are not impacted by mutant Ctnnb1. First, we generated $Atoh1\text{-}Cre^{+/-}$;$Ctnnb1^{+/lox(ex3)}$ mice because *Atoh1-Cre* drives efficient recombination in GNPCs, generating medulloblastomas in conditional *Ptch1* mice (see Supplementary Fig. 8a–j and ref. 7). We also used the *Atoh1* enhancer element present in the *Atoh1-Cre* allele to drive expression of a constitutively active Ctnnb1-green fluorescence fusion protein (GFP) in GNPCs ($Atoh1\text{-}Ctnnb1^{\Delta N90GFP}$, Supplementary Fig. 8k–o)[22]. Neither $Atoh1\text{-}Cre^{+/-}$;$Ctnnb1^{+/lox(ex3)}$ nor $Atoh1\text{-}Ctnnb1^{\Delta N90GFP}$ mice (more than 20 mice examined each) developed hyperplasia or masses within the URL or EGL. Concordantly, aberrant Ctnnb1 signalling did not impact the proliferation of GNPCs *ex vivo* (Supplementary Fig. 8p). Thus, in contrast to aberrant Shh signalling, mutant Ctnnb1 does not appear to disrupt cell-cycle or differentiation control in GNPCs.

In stark contrast to the cerebellum, by E16.5 all *Ctnnb1*-mutant mice developed aberrant cell collections in the dorsal brainstem that persisted into adulthood (exact Mann–Whitney $P < 0.005$, Fig. 2a–f). These cells were marked by Olig3 and Pax6, which suggested they may be derived from progenitor cells within the LRL[19,23] (Fig. 2d, e). This abnormality was independent of *Tp53* status and did not involve the floor plate that is not targeted by *Blbp-Cre* (Supplementary Fig. 9). Progenitors within the embryonic dorsal brainstem proliferate to produce daughter cells that express specific marker proteins and follow complex migration streams to their respective nuclei in the developing brainstem (Supplementary Fig. 4)[15]. We observed no significant differences in the overall proliferation (Ki67 labelling), apoptosis (TdT-mediated dUTP nick end labelling) or cell-cycle duration (5-bromo-2-deoxyuridine pulse-chase) of progenitors in the dorsal brainstem of *Ctnnb1*-mutant versus control mice (Fig. 2c, data not shown). However, a significant fraction of proliferating cells within *Ctnnb1*-mutant dorsal brainstems expressed Zic1 (37% Zic1^{+}/Ki67^{+} = 122/322; Fig. 2c, f–h). This expression is aberrant because Zic1 normally marks postmitotic mossy-fibre neuron precursors as they exit the dorsal brainstem to form nuclei in the ventral brainstem[23,24] (Fig. 2g). Thus mutant Ctnnb1 might stall the dorso-ventral migration of brainstem neuron precursors, resulting in aberrant dorsal cell collections[25]. To test this, we used *in utero* GFP electroporation to track the fate of embryonic dorsal brainstem precursors (Fig. 2i–q and Supplementary Figs 10 and 11). GFP-labelled Zic1^{+} mossy-fibre neuron precursors underwent normal migration from the dorsal brainstem to the pontine grey nucleus and other brainstem nuclei in control mice (Fig. 2k–n and Supplementary Fig. 11). In contrast, mutation of *Ctnnb1* markedly reduced the numbers of precursors transiting from the dorsal brainstem to the pontine grey nucleus (Fig. 2o–q; exact Mann–Whitney, $P < 0.05$). Together, these data demonstrate that mutant Ctnnb1 disrupts the normal differentiation and migration of progenitor cells on the dorsal brainstem, resulting in the accumulation of aberrant cell collections. These cells may include stalled mossy-fibre neuron precursors, but further work is required to determine their precise lineage.

Aberrant cell collections in the dorsal brainstem of *Ctnnb1*-mutant mice are reminiscent of the EGL hyperplasia that precedes formation of SHH-subtype medulloblastoma in the cerebellum of *Ptch1*-deficient mice[26]. Therefore we aged *Ctnnb1*-mutant mice harbouring $Tp53^{+/+}$, $Tp53^{+/flx}$ or $Tp53^{flx/flx}$ alleles to test if WNT-subtype medulloblastomas might arise from the dorsal brainstem (n more than 54 mice per genotype). Aberrant cell collections persisted throughout adulthood on the dorsal brainstem of all *Ctnnb1*-mutant;$Tp53^{+/+}$ mice, but these animals did not develop medulloblastoma or tumours in any part of the hindbrain (median follow up 365 days). In contrast, 2 out of 10 *Ctnnb1*-mutant;$Tp53^{flx/flx}$ mice aged older than 6 months harboured asymptomatic tumours that were confined to the dorsal brainstem (Supplementary Fig. 12). When aged for longer periods, 15% ($n = 8/55$) of *Ctnnb1*-mutant;$Tp53^{flx/flx}$ and 4% ($n = 2/54$) *Ctnnb1*-mutant;$Tp53^{+/flx}$

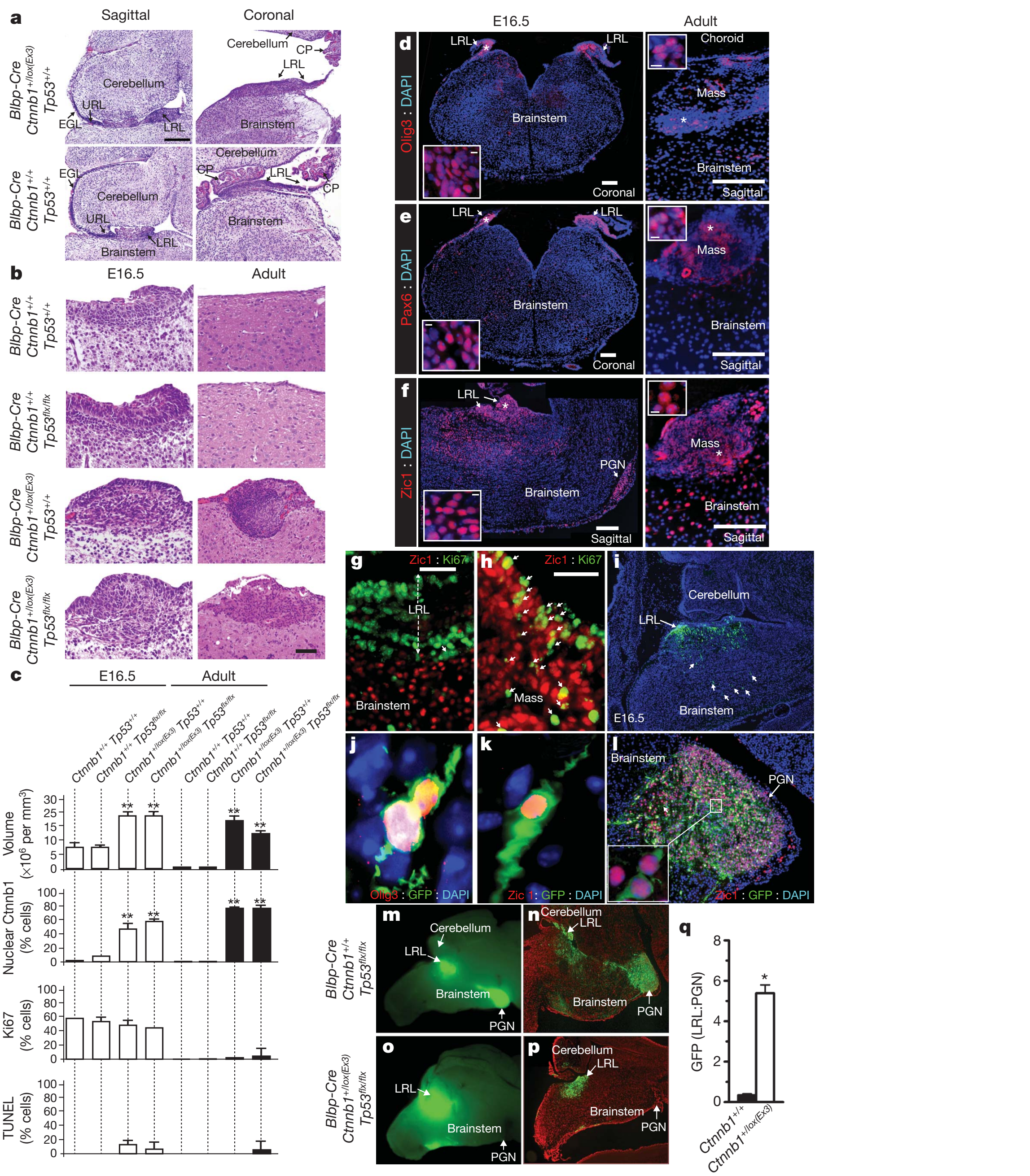


Figure 2 | Mutant-*Ctnnb1* causes aberrant accumulation of LRL cells. **a**, Low- (scale bar, 180 μm) and **b**, high- (scale bar, 50 μm) power views of LRL/dorsal brainstem in *Ctnnb1* mutant and wild-type E16.5 embryos; **b** includes the corresponding adult brainstem region. **c**, Volume and indicated immunoreactivity differences between *Ctnnb1*-mutant and wild-type LRL ($n \geq 3$ mice per group; bars, mean ± s.d.). Immunofluorescence of Olig3 (**d**), Pax6 (**e**) and Zic1 (**f**) in *Ctnnb1*-mutant E16.5 LRL (left) and aberrant adult dorsal brainstem masses (right) (scale bar, 180 μm). Inset, high-power views of '*' (scale bar, 5 μm). **g**, Postmitotic mossy-fibre precursor neurons ($Zic1^+/Ki67^-$) exit the proliferating E16.5 control LRL. **h**, *Ctnnb1*-mutant LRL contains aberrant proliferating $Zic1^+$ precursors (arrows; scale bar, 50 μm). **i**, GFP-electroporated wild-type LRL marks $Olig3^+$ cells (**j**) and migrating precursors (arrows in **i**) that include $Zic1^+$ mossy-fibre neurons (**k**) that form the pontine grey nucleus (**l**). GFP-fluorescence of whole (**m**, **o**) and sectioned (**n**, **p**) *Ctnnb1*-mutant and wild-type P0 hindbrains electroporated at E12.5. **q**, Mean ± s.d. of LRL/pontine grey nucleus GFP fluorescence in whole hindbrains of three *BlBp-Cre*;$Ctnnb1^{+/+}$ and five *Blbp-Cre*;$Ctnnb1^{+/lox(Ex3)}$ mice (graphs; $*P \leq 0.05$, $**P \leq 0.005$, exact Mann–Whitney).

mice developed 'classic' medulloblastomas that were $Zic1^+$ and contained populations of nuclear-$Ctnnb1^+/Olig3^+$ cells (median follow up 290 and 287 days, respectively; Fig. 3a–d). These mouse medulloblastomas displayed an immunoprofile similar to human WNT-subtype tumours and were invariably connected with the brainstem (Fig. 3d, e and Supplementary Fig. 13). In contrast, mouse models of human SHH-subtype medulloblastoma[21,27,28] are nuclear-Ctnnb1 negative, arise within the cerebellum and do not invade the brainstem (Fig. 3d, e). Together, these data support the hypothesis that progenitor cells within the dorsal brainstem are susceptible to transformation by concurrent mutation in *Ctnnb1* and *Tp53*, resulting in the formation of tumours that mimic the anatomical features of human WNT-subtype medulloblastoma. Deletion of *Tp53* is presumably required to allow key second mutations during transformation of the LRL in *Ctnnb1*-mutant mice. Notably, we have observed two cases of *TP53*-mutant human WNT-subtype medulloblastoma, suggesting this gene also suppresses these tumours in humans (Supplementary Fig. 14).

To test further the fidelity of *Ctnnb1*-mutant mouse medulloblastoma as a model of human WNT-subtype disease, we compared the tumour transcriptomes in the two species using an algorithm we have

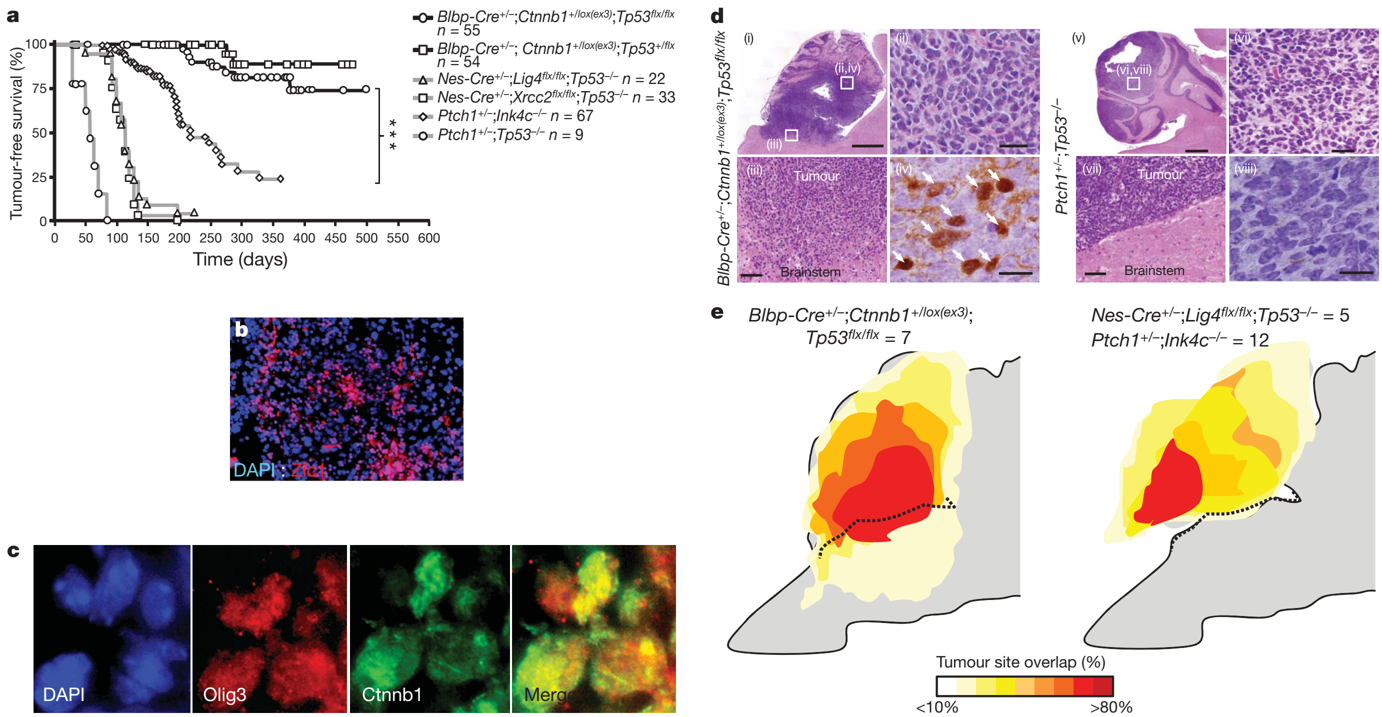


Figure 3 | Mutant-*Ctnnb1* and SHH-subtype mouse medulloblastomas are anatomically distinct. **a**, Tumour-free survival of SHH-subtype medulloblastoma mouse models (*Nes-Cre*$^{+/-}$;*Lig4*$^{flx/flx}$;*Tp53*$^{-/-}$, *Nes-Cre*$^{+/-}$;*Xrcc2*$^{flx/flx}$;*Tp53*$^{-/-}$, *Ptch1*$^{+/-}$;*Ink4c*$^{-/-}$, *Ptch1*$^{+/-}$;*Tp53*$^{-/-}$, data from refs 14, 27, 28) and *Ctnnb1*-mutant;*Tp53*$^{flx/flx}$ and *Ctnnb1*-mutant;*Tp53*$^{+/flx}$ mice. ***Log rank $P < 0.0001$. Immunofluorescence of (**b**) Zic1 and (**c**) Olig3 and Ctnnb1 expression in a *Ctnnb1*-mutant;*Tp53*$^{flx/flx}$ medulloblastoma. **d**, Haematoxylin and eosin-stained low- (i, v; scale bar, 800 μm) and high- (ii, vi; scale bar, 25 μm) power views of mouse medulloblastomas and tumour–brainstem interface (iii, vii; scale bar, 50 μm). *Ctnnb1* immunostaining (iv, viii; scale bar, 10 μm, arrows indicate nuclear immunoreactivity). Boxes indicate location of high-power views. **e**, Frequency and anatomical site of mouse medulloblastomas.

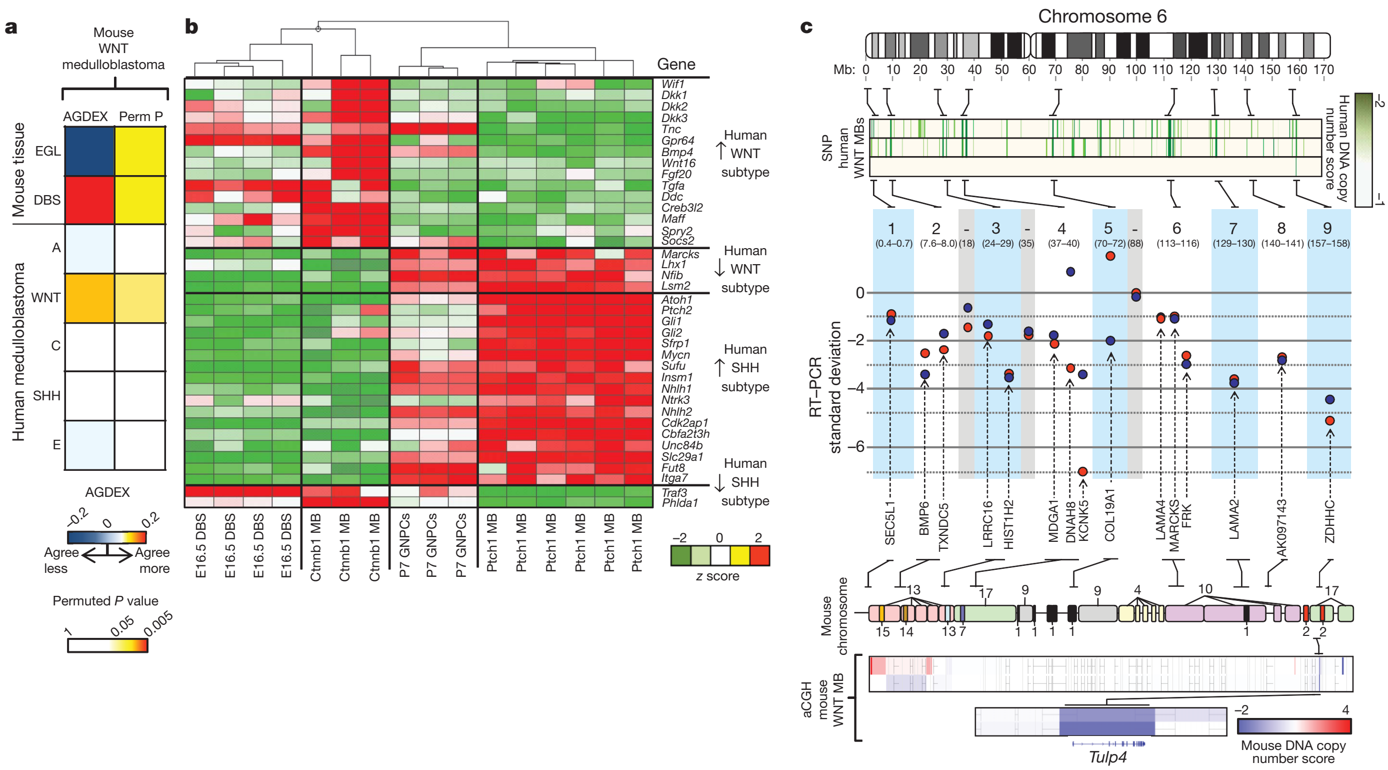


Figure 4 | Mutant-*Ctnnb1* mouse medulloblastomas recapitulate the molecular characteristics of human WNT-subtype disease. **a**, AGDEX comparison of *Ctnnb1*-mutant;*Tp53*$^{flx/flx}$ mouse medulloblastoma, and mouse EGL, E16.5 dorsal brainstem (DBS) and human medulloblastoma subgroups. **b**, Unsupervised clustering of human WNT- and SHH-subtype medulloblastoma signature orthologue expression in E16.5 DBS, *Ctnnb1*-mutant;*Tp53*$^{flx/flx}$ mouse medulloblastoma (Ctnnb1 MB), P7 GNPCs and *Ptch1*$^{+/-}$;*Tp53*$^{-/-}$ medulloblastoma (Ptch1 MB). **c**, Top–bottom: nine single nucleotide polymorphism-inferred homozygous deletions in three human WNT-subtype medulloblastomas; real-time PCR validation of deletions in the human tumours (s.d. below the mean diploid copy number); mouse chromosomal regions syntenic for human chromosome 6; array comparative genomic hybridization-inferred copy number in *Ctnnb1*-mutant;*Tp53*$^{flx/flx}$ mouse medulloblastomas identifies common syntenic deletion of *TULP4*.

developed for cross-species genomic comparisons[13]. Remarkably, the transcriptome ($n = 11{,}049$ orthologues) of *Ctnnb1*-mutant;*Tp53*$^{flx/flx}$ medulloblastomas matched only human WNT-subtype medulloblastoma and the cells of the embryonic dorsal brainstem (both permuted $P < 0.05$), validating it as a model of this human tumour subtype and further pinpointing the brainstem as the source of WNT-subtype medulloblastomas (Fig. 4a, b). Finally, because human WNT-subtype medulloblastomas selectively delete chromosome 6 (ref. 3), we looked in *Ctnnb1*-mutant mouse medulloblastomas to see if syntenic regions of this chromosome are deleted (Fig. 4c). DNA microarray analysis identified a single common deletion of mouse chromosome 17 3.2 cM/human 6q25.3 in tumours in the two species. This locus encodes a single gene, *TULP4*, that is a distant member of the tubby-gene family implicated in regulating neuronal cell apoptosis[29]. Thus *Ctnnb1*-mutant;*Tp53*$^{flx/flx}$ mouse medulloblastomas accurately model the molecular characteristics of human WNT-subtype tumours and pinpoint *TULP4* as a novel candidate suppressor gene of these tumours. By demonstrating that subtypes of medulloblastoma have distinct cellular origins, our data should significantly accelerate the hunt for curative treatments of these diseases, which must now account for the different developmental origins of these tumours.

METHODS SUMMARY

MRI analysis. MRI images of patients were spatially normalized into a standard stereotaxic space for quantitative comparison of tumour location (SPM5; www.fil.ion.ucl.ac.uk/spm). Radiologists masked to patient subtype determined the three-dimensional location of the tumour or surgical cavity relative to pre-defined anatomical landmarks.

Expression mapping. The expression of mouse orthologues of key signature genes of human WNT- and SHH-subtype medulloblastoma (Supplementary Data Set 1 and Supplementary Table) were mapped in the developing mouse hindbrain using four publically accessible data sets (see Supplementary Methods).

Mouse studies. *Blbp-Cre*, *Ctnnb1*$^{lox(ex3)/lox(ex3)}$, *Atoh1-Cre* and *Tp53*$^{flx/flx}$ mice were bred to generate appropriate genotypes and subjected to clinical surveillance for signs of tumour development. *RosaYFP* and *RosaLacz* reporter strains traced the lineage of Cre-recombined cells. Mouse tumours comprised at least 85% tumour cells. *Atoh1-Ctnnb1*$^{\Delta N90}$ transgenic mice were generated by pro-nuclear injection. *In utero* electroporation and cell tracking were performed by anaesthetizing pregnant mice of the appropriate genotype. The uterus was externalized and the dorsal brainstem of E12.5 embryos electroporated with *CMV-eGFP* plasmid DNA. GNPCs for culture studies were isolated from postnatal day 7 *Atoh1-GFP* transgenic mice. GFP^{+} cells (2×10^{5} per well) were cultured in poly-D-lysine-coated 96-well plates and challenged with mutant-*Ctnnb1-GFP*, control *GFP* virus, Wnt1 protein ($50\,\mathrm{ng\,ml^{-1}}$) or Shh supernatant ($3\,\mu\mathrm{g\,ml^{-1}}$) before pulsing with [methyl-3H]thymidine and scintillation counting.

Histology, messenger RNA and DNA microarray profiling. Immunohistochemistry was performed using routine techniques and primary antibodies of the appropriate tissues as described (Supplementary Methods). Cells undergoing apoptosis were detected with the Apoptag kit (Millipore, S7100). Messenger RNA expression (GEO accession number GSE24628) and DNA copy number profiles (available at http://stjuderesearch.org/site/authors/gilbertson) were generated from mouse and human tissues using appropriate microarray platforms as detailed (Supplementary Methods). Reverse transcriptase real-time PCR and gene re-sequencing of human medulloblastomas were performed as described previously[3]. Messenger RNA expression and DNA microarray profiles of human and mouse medulloblastomas were integrated using established and novel bioinformatic and statistical approaches.

Received 26 September 2009; accepted 15 October 2010.
Published online 8 December 2010.

1. Ellison, D. W. *et al.* β-Catenin status predicts a favorable outcome in childhood medulloblastoma: the United Kingdom Children's Cancer Study Group Brain Tumour Committee. *J. Clin. Oncol.* **23,** 7951–7957 (2005).
2. Gajjar, A. *et al.* Risk-adapted craniospinal radiotherapy followed by high-dose chemotherapy and stem-cell rescue in children with newly diagnosed medulloblastoma (St Jude Medulloblastoma-96): long-term results from a prospective, multicentre trial. *Lancet Oncol.* **7,** 813–820 (2006).
3. Thompson, M. C. *et al.* Genomics identifies medulloblastoma subgroups that are enriched for specific genetic alterations. *J. Clin. Oncol.* **24,** 1924–1931 (2006).
4. Kool, M. *et al.* Integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features. *PLoS ONE* **3,** e3088 (2008).
5. Gilbertson, R. J. & Ellison, D. W. The origins of medulloblastoma subtypes. *Annu. Rev. Pathol.* **3,** 341–365 (2008).
6. Goodrich, L. V., Milenkovic, L., Higgins, K. M. & Scott, M. P. Altered neural cell fates and medulloblastoma in mouse patched mutants. *Science* **277,** 1109–1113 (1997).
7. Schuller, U. *et al.* Acquisition of granule neuron precursor identity is a critical determinant of progenitor cell competence to form Shh-induced medulloblastoma. *Cancer Cell* **14,** 123–134 (2008).
8. Yang, Z. J. *et al.* Medulloblastoma can be initiated by deletion of Patched in lineage-restricted progenitors or stem cells. *Cancer Cell* **14,** 135–145 (2008).
9. Romer, J. T. *et al.* Suppression of the Shh pathway using a small molecule inhibitor eliminates medulloblastoma in Ptc1$^{+/-}$p53$^{-/-}$ mice. *Cancer Cell* **6,** 229–240 (2004).
10. Rudin, C. M. *et al.* Treatment of medulloblastoma with Hedgehog pathway inhibitor GDC-0449. *N. Engl. J. Med.* **2,** 1173–1178 (2009).
11. Louis, D., Ohgaki, H., Wiestler, O. & Cavenee, W. (eds) *World Health Organization Classification of Tumours of the Central Nervous System* (International Agency for Research on Cancer, 2007).
12. Huang, X. *et al.* Transventricular delivery of Sonic hedgehog is essential to cerebellar ventricular zone development. *Proc Natl Acad. Sci. USA* **107,** 8422–8427 (2010).
13. Johnson, R. A. *et al.* Cross-species genomics matches driver mutations and cell compartments to model ependymoma. *Nature* **466,** 632–636 (2010).
14. Lee, Y. *et al.* A molecular fingerprint for medulloblastoma. *Cancer Res.* **63,** 5428–5437 (2003).
15. Ray, R. S. & Dymecki, S. M. Rautenlippe Redux—toward a unified view of the precerebellar rhombic lip. *Curr. Opin. Cell Biol.* **21,** 741–747 (2009).
16. Morales, D. & Hatten, M. E. Molecular markers of neuronal progenitors in the embryonic cerebellar anlage. *J. Neurosci.* **26,** 12226–12236 (2006).
17. Harada, N. *et al.* Intestinal polyposis in mice with a dominant stable mutation of the beta-catenin gene. *EMBO J.* **18,** 5931–5942 (1999).
18. Hegedus, B. *et al.* Neurofibromatosis-1 regulates neuronal and glial cell differentiation from neuroglial progenitors *in vivo* by both cAMP- and Ras-dependent mechanisms. *Cell Stem Cell* **1,** 443–457 (2007).
19. Storm, R. *et al.* The bHLH transcription factor Olig3 marks the dorsal neuroepithelium of the hindbrain and is essential for the development of brainstem nuclei. *Development* **136,** 295–305 (2009).
20. Jonkers, J. *et al.* Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. *Nature Genet.* **29,** 418–425 (2001).
21. Wetmore, C., Eberhart, D. E. & Curran, T. Loss of p53 but not ARF accelerates medulloblastoma in mice heterozygous for patched. *Cancer Res.* **61,** 513–516 (2001).
22. Chenn, A. & Walsh, C. A. Regulation of cerebral cortical size by control of cell cycle exit in neural precursors. *Science* **297,** 365–369 (2002).
23. Landsberg, R. L. *et al.* Hindbrain rhombic lip is comprised of discrete progenitor cell populations allocated by Pax6. *Neuron* **48,** 933–947 (2005).
24. DiPietrantonio, H. J. & Dymecki, S. M. Zic1 levels regulate mossy fiber neuron position and axon laterality choice in the ventral brain stem. *Neuroscience* **162,** 560–573 (2009).
25. Farago, A. F., Awatramani, R. B. & Dymecki, S. M. Assembly of the brainstem cochlear nuclear complex is revealed by intersectional and subtractive genetic fate maps. *Neuron* **50,** 205–218 (2006).
26. Oliver, T. G. *et al.* Loss of patched and disruption of granule cell development in a pre-neoplastic stage of medulloblastoma. *Development* **132,** 2425–2439 (2005).
27. Uziel, T. *et al.* The tumor suppressors Ink4c and p53 collaborate independently with Patched to suppress medulloblastoma formation. *Genes Dev.* **19,** 2656–2667 (2005).
28. Frappart, P. O. *et al.* Recurrent genomic alterations characterize medulloblastoma arising from DNA double-strand break repair deficiency. *Proc. Natl Acad. Sci. USA* **106,** 1880–1885 (2009).
29. Ikeda, A., Ikeda, S., Gridley, T., Nishina, P. M. & Naggert, J. K. Neural tube defects and neuroepithelial cell death in Tulp3 knockout mice. *Hum. Mol. Genet.* **10,** 1325–1334 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements R.J.G. holds the Howard C. Schott Research Chair from the Malia's Cord Foundation, and is supported by grants from the National Institutes of Health (R01CA129541, P01CA96832 and P30CA021765), the Collaborative Ependymoma Research Network and by the American Lebanese Syrian Associated Charities. We are grateful to A. Chenn, J. Johnson and C. Birchmeier for their gifts of reagents and the staff of the Hartwell Center for Bioinformatics and Biotechnology and ARC at St Jude Children's Research Hospital for technical assistance.

Author Contributions R.J.G. conceived the research and planned experiments. P.G. also planned and conducted most of the experiments. Y.T., G.R., D.S.C., M.C.T., T.H., H.P., J.M., J.C.L., Y.L., F.Z., C.E., S.C.C., M.F.R., P.J.M. and R.W.-R. conducted experiments. D.F. and S.P. provided bioinformatic expertise. A.G., F.A.B. and R.A.S. provided clinical advice and tumour samples. D.H.G. provided the *Blbp-Cre* mouse and data. M.M.T. provided the *Ctnnb1*$^{lox(ex3)/lox(ex3)}$ mouse. Z.P. and R.O. reviewed and analysed the human MRI scans. D.W.E. provided pathology review. All authors contributed to writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.J.G. (Richard.Gilbertson@stjude.org).

doi:10.1038/nature09588

A selective role for dopamine in stimulus–reward learning

Shelly B. Flagel[1]*, Jeremy J. Clark[2]*, Terry E. Robinson[3], Leah Mayo[1], Alayna Czuj[3], Ingo Willuhn[2], Christina A. Akers[2], Sarah M. Clinton[1], Paul E. M. Phillips[2] & Huda Akil[1]

Individuals make choices and prioritize goals using complex processes that assign value to rewards and associated stimuli. During Pavlovian learning, previously neutral stimuli that predict rewards can acquire motivational properties, becoming attractive and desirable incentive stimuli. However, whether a cue acts solely as a predictor of reward, or also serves as an incentive stimulus, differs between individuals. Thus, individuals vary in the degree to which cues bias choice and potentially promote maladaptive behaviour. Here we use rats that differ in the incentive motivational properties they attribute to food cues to probe the role of the neurotransmitter dopamine in stimulus–reward learning. We show that intact dopamine transmission is not required for all forms of learning in which reward cues become effective predictors. Rather, dopamine acts selectively in a form of stimulus–reward learning in which incentive salience is assigned to reward cues. In individuals with a propensity for this form of learning, reward cues come to powerfully motivate and control behaviour. This work provides insight into the neurobiology of a form of stimulus–reward learning that confers increased susceptibility to disorders of impulse control.

Dopamine is central for reward-related processes[1,2], but the exact nature of its role remains controversial. Phasic neurotransmission in the mesolimbic dopamine system is initially triggered by the receipt of reward (unconditional stimulus, US), but shifts to a cue that predicts a reward (conditional stimulus, CS) after associative learning[3,4]. Dopamine responsiveness appears to encode discrepancies between rewards received and those predicted, consistent with a 'prediction error' teaching signal used in formal models of reinforcement learning[5,6]. Therefore, a popular hypothesis is that dopamine is used to update the predictive value of stimuli during associative learning[7]. In contrast, others have argued that the role of dopamine in reward is in attributing Pavlovian incentive value to cues that signal reward, rendering them desirable in their own right[8–11], and thereby increasing the pool of positive stimuli that have motivational control over behaviour. Until now it has been difficult to determine whether dopamine mediates the predictive or the motivational properties of reward-associated cues, because these two features are often acquired together. However, the extent to which a predictor of reward acquires incentive value differs between individuals, providing the opportunity to parse the role of dopamine in stimulus–reward learning.

Individual variation in behavioural responses to reward-associated stimuli can be seen using one of the simplest reward paradigms, Pavlovian conditioning. If a CS is presented immediately before US delivery at a separate location, some animals approach and engage the CS itself and go to the location of food delivery only upon CS termination. This conditional response (CR), which is maintained by Pavlovian contingency[12], is called 'sign-tracking' because animals are attracted to the cue or sign that indicates impending reward delivery. However, other individuals do not approach the CS, but during its presentation engage the location of US delivery, even though the US is not available until CS termination. This CR is called 'goal-tracking'[13]. The CS is an effective predictor in animals that learn either a sign-tracking or a goal-tracking response; it acts as an excitor, evoking a CR in both. However, only in sign-trackers is the CS an attractive incentive stimulus, and only in sign-trackers is it strongly desired (that is, 'wanted'), in the sense that animals will work avidly to get it[14]. In rats selectively bred for differences in locomotor responses to a novel environment[15], high responders to novelty (bHR rats) consistently learn a sign-tracking CR but low responders to novelty (bLR rats) consistently learn a goal-tracking CR[16]. Here, we exploit these predictable phenotypes in the selectively bred rats, as well as normal variation in outbred rats, to probe the role of dopamine transmission in stimulus–reward learning in individuals that vary in the incentive value they assign to reward cues.

Stimulus–reward learning

bHR and bLR rats from the twentieth generation of selective breeding (S20) were used for behavioural analysis of Pavlovian conditional approach behaviour[16] (Fig. 1a–e). When presentation of a lever-CS was paired with food delivery both bHR and bLR rats developed a Pavlovian CR, but as we have described previously[16], the topography of the CR was different in the two groups. With training, bHR rats came to rapidly approach and engage the lever-CS (Fig. 1a, b), whereas upon CS presentation bLR rats came to rapidly approach and engage the location where food would be delivered (Fig. 1c and d; see detailed statistics in Supplementary Information). Both bHR and bLR rats acquired their respective CRs as a function of training, given that there was a significant effect of number of sessions for all measures of sign-tracking behaviour for bHR rats (Fig. 1a, b; $P \leq 0.0001$), and of goal-tracking behaviour for bLR rats (Fig. 1c, d; $P \leq 0.0001$). Furthermore, bHR and bLR rats learned their respective CRs at the same rate, as indicated by analyses of variance in which session was treated as a continuous variable and the phenotypes were directly compared. There were non-significant phenotype × session interactions for (1) the number of contacts with the lever-CS for bHR rats versus the food-tray for bLR rats ($F_{(1, 236)} = 3.02$, $P = 0.08$) and (2) the latency to approach the lever-CS for bHR rats versus the

[1]Molecular and Behavioral Neuroscience Institute, University of Michigan, Michigan, USA. [2]Department of Psychiatry and Behavioral Sciences and Department of Pharmacology, University of Washington, Washington, USA. [3]Department of Psychology, University of Michigan, Michigan, USA.
*These authors contributed equally to this manuscript.

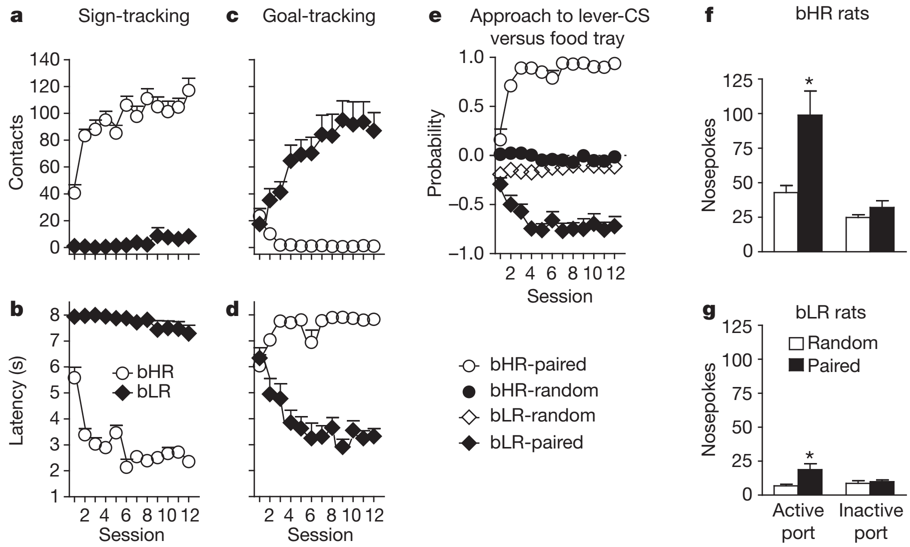


Figure 1 | Development of sign-tracking versus goal-tracking CRs in bHR and bLR rats. Behaviour directed towards the lever-CS (sign-tracking) is shown in **a** and **b** and behaviour directed towards the food-tray (goal-tracking) is shown in **c** and **d** ($n = 10$ per group). Data are shown as mean + s.e.m. **a**, Number of lever-CS contacts made during the 8-s CS period. **b**, Latency to the first lever-CS contact. **c**, Number of food-tray beam breaks during lever-CS presentation. **d**, Latency to the first beam break in the food-tray during lever-CS presentation. For all of these measures (**a**–**d**) there was a significant effect of phenotype, session, and a phenotype × session interaction ($P \leq 0.0001$). **e**, Probability of approach to the lever minus the probability of approach to the food-tray shown as mean+ s.e.m. A score of zero indicates that neither approach to the lever-CS nor approach to the food-tray was dominant. **f**, **g**, Test for conditioned reinforcement illustrated as the number (mean + s.e.m.) of active and inactive nosepokes in bred rats that received either paired (bHR rats, $n = 10$; bLR rats, $n = 9$) or pseudorandom (bHR rats, $n = 9$; bLR rats, $n = 9$) CS–US presentations. Rats in the paired groups poked more in the active port than did random groups of the same phenotype (*$P < 0.02$), but the magnitude of this effect was greater for bHR rats (phenotype × group interaction, $P = 0.04$).

food-tray for bLR rats ($F_{(1,\ 236)} = 0.93$, $P = 0.34$). Importantly, rats that received non-contiguous (pseudorandom) presentations of the CS and the US did not learn either a sign-tracking or a goal-tracking CR (Fig. 1e).

These data indicate that the CS acquired one defining property of an incentive stimulus in bHR rats but not bLR rats: the ability to attract. Another feature of an incentive stimulus is to be 'wanted' and as such animals should work to obtain it[10,17]. Therefore, we quantified the ability of the lever-CS to serve as a conditioned reinforcer in the two groups (Fig. 1f, g) in the absence of the food-US. Following Pavlovian training, rats were given the opportunity to perform an instrumental response (a nosepoke) for presentation of the lever-CS. Responses into a port designated 'active' resulted in the brief presentation of the lever-CS and responses into an 'inactive' port were without consequence. Both conditioned bHR and bLR rats made more active than inactive nose pokes, and more active nose pokes than control groups that received pseudorandom presentations of the CS and the US (Fig. 1f, g; detailed statistics in Supplementary Information). However, the lever-CS was a more effective conditioned reinforcer in bHR rats than in bLR rats, as indicated by a significant phenotype × group interaction for active nose pokes ($F_{(1, 33)} = 4.82$, $P = 0.04$), which controls for basal differences in nose-poke responding. Moreover, in outbred rats in which this baseline difference in responding does not exist, we have found similar results, indicating that the lever-CS is a more effective conditioned reinforcer for sign-trackers than goal-trackers[14]. In summary, the lever-CS was equally predictive, evoking a CR in both groups, but it acquired two properties of an incentive stimulus to a greater degree in bHR rats than bLR rats: it was more attractive, as indicated by approach behaviour (Fig. 1a) and more desirable, as indicated by its ability to serve as a conditioned reinforcer (Fig. 1f, g).

Dopamine signalling during stimulus–reward learning

The core of the nucleus accumbens is an important anatomical substrate for motivated behaviour[18,19] and has been specifically implicated as a site where dopamine acts to mediate the acquisition and/or performance of Pavlovian conditional approach behaviour[20–23]. Therefore, we used fast-scan cyclic voltammetry (FSCV) at carbon-fibre microelectrodes[24] to characterize the pattern of phasic dopamine signalling in this region during Pavlovian conditioning (see Supplementary Fig. 1 for recording locations). Similarly to surgically naive animals, bHR rats learned a sign-tracking CR (session effect on lever contacts: $F_{(5,\ 20)} = 5.76$, $P = 0.002$) and bLR rats learned a goal-tracking CR (session effect on food-receptacle contacts: $F_{(5,\ 20)} = 5.18$, $P = 0.003$) during neurochemical data collection (Supplementary Fig. 2). Changes in latency during learning were very similar in each group for their respective CRs (main effect of session: $F_{(5,\ 40)} = 10.5$, $P < 0.0001$; main effect of phenotype: $F_{(1,\ 8)} = 0.13$, $P = 0.73$; session × phenotype interaction: $F_{(5, 40)} = 1.16$, $P = 0.35$), indicating that the CS acts as an equivalent predictor of reward in both groups. Therefore, if CS-evoked dopamine release encodes the strength of the reward prediction, as previously postulated[5–7], it should increase to a similar degree in both groups during learning; however, if it encodes the attribution of incentive value to the CS, then it should increase to a greater degree in sign-trackers than in goal-trackers. During the acquisition of conditional approach, CS-evoked dopamine release (Fig. 2 and Supplementary Fig. 3) increased in bHR rats relative to unpaired controls (pairing × session interaction: $F_{(5,\ 35)} = 4.58$, $P = 0.003$), but there was no such effect in bLR rats (Fig. 2 and Supplementary Fig. 3; pairing × session interaction: $F_{(5,\ 35)} = 0.94$, $P = 0.46$). Indeed, the trial-by-trial correlation between CS-evoked dopamine release and trial number was significant for bHR rats ($r^2 = 0.14$, $P < 0.0001$) but not bLR rats ($r^2 = 0.003$, $P = 0.54$), producing significantly different slopes ($P = 0.005$) and higher CS-evoked dopamine release in bHR rats after acquisition (Supplementary Fig. 4, session 6; $P = 0.04$). US-evoked dopamine release also differed between bHR and bLR rats during training (session × phenotype interaction: $F_{(5,\ 40)} = 6.09$, $P = 0.0003$), but for this stimulus dopamine release was lower after acquisition in bHR rats (session 6; $P = 0.002$; Supplementary Fig. 4). Collectively, these data highlight that bHR and bLR rats produce fundamentally different patterns of dopamine release in response to reward-related stimuli during learning (see Supplementary Videos 1 and 2). The CS and US signals diverge in bHR rats (stimulus × session interaction: $F_{(5,\ 40)} = 5.47$,

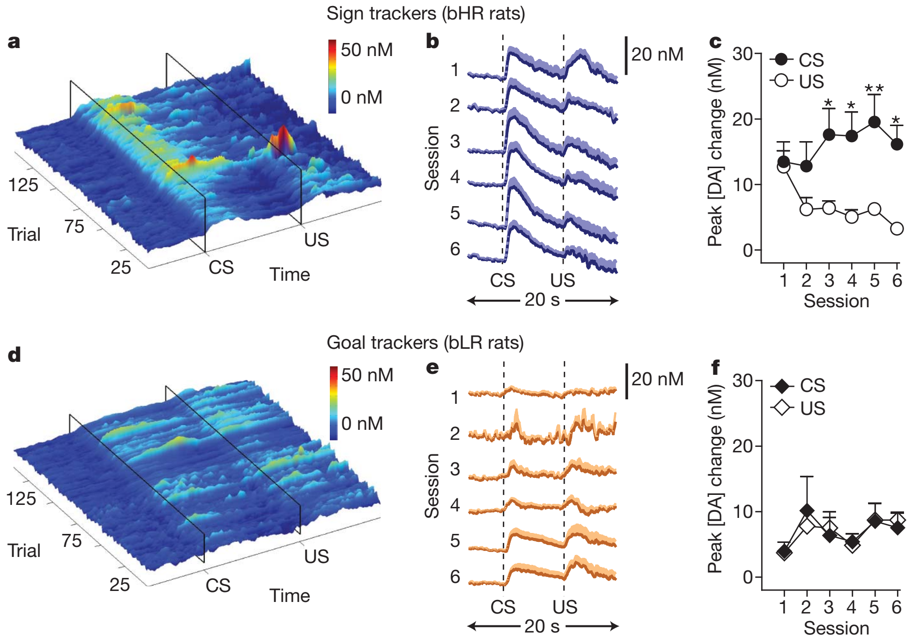


Figure 2 | Phasic dopamine signalling in response to CS and US presentation during the acquisition of Pavlovian conditional approach behaviour in bHR and bLR rats. Phasic dopamine release was recorded in the core of the nucleus accumbens using FSCV over six days of training. **a**, **d**, Representative surface plots depict trial-by-trial fluctuations in dopamine concentration [DA] during the twenty-second period around CS and US presentation in individual animals throughout training. **b**, **e**, Change in dopamine concentration (mean + s.e.m.) in response to CS and US presentation for each session of conditioning. **c**, **f**, Change in peak amplitude (mean + s.e.m.) of the dopamine signal observed in response to CS and US presentation for each session of conditioning ($n = 5$ per group; Bonferroni post-hoc comparison between CS- and US-evoked dopamine release: $*P < 0.05$; $**P < 0.01$). Panels **a**–**c** demonstrate that bHR rats, which developed a sign-tracking CR, show increasing phasic dopamine responses to CS presentation and decreasing responses to US presentation across the six sessions of training. In contrast, panels **d**–**f** demonstrate that bLR rats, which developed a goal-tracking CR, maintain phasic responses to US presentation throughout training.

$P = 0.0006$; Fig. 2c) but not bLR rats (stimulus × session interaction: $F_{(5, 40)} = 0.28$, $P = 0.92$; Fig. 2f).

Importantly, experiments conducted in commercially obtained outbred rats reproduced the pattern of dopamine release observed in the selectively bred rats (Fig. 3 and Supplementary Fig. 4). Specifically, there was an increase in CS-evoked and a decrease in US-evoked dopamine release during learning in outbred rats that learned a sign-tracking CR (stimulus × session interaction: $F_{(5, 50)} = 4.43$, $P = 0.002$; Fig. 3d), but not in those that learned a goal-tracking CR (stimulus × session interaction: $F_{(5, 40)} = 0.48$, $P = 0.72$; Fig. 3f). To test the robustness of these patterns of dopamine release, a subset of outbred rats received extended training. During four additional sessions, the profound differences in dopamine release between sign- and goal-trackers were stable (Supplementary Fig. 5), demonstrating that these differences are not limited to the initial stages of learning. The consistency of these dopamine patterns in selectively bred and outbred rats indicates that they are neurochemical signatures for sign- and goal-trackers rather than an artefact of selective breeding.

Stimulus–reward learning under dopamine blockade

Given the disparate patterns of dopamine signalling observed during learning a sign- versus goal-tracking CR, we tested whether the acquisition and performance of these CRs were differentially dependent on dopamine transmission. Systemic administration of flupenthixol, a nonspecific dopamine receptor antagonist, attenuated performance of the CR for both bHR and bLR rats. This effect was clearly evident when the antagonist was administered during training (Fig. 4, sessions 1–7). It was also observed after the rats had already acquired their respective CR (Supplementary Fig. 6), but this latter finding needs to be interpreted cautiously because of a non-specific effect on activity (Supplementary Fig. 6e). More importantly, when examined off flupenthixol during the eighth test session, bHR rats still failed to demonstrate a sign-tracking CR ($P \leq 0.01$ versus saline, session 8; Fig. 4a–c), indicating that dopamine is necessary for both the performance and the learning of a sign-tracking CR, consistent with previous findings[21]. In contrast, flupenthixol had no effect on learning the CS–US association that lead to a goal-tracking CR ($P \geq 0.6$ versus saline, session 8; Fig. 4d–f), because on the drug-free session bLR rats showed a fully developed goal-tracking CR—their session 8 performance differed significantly from their session 1 performance ($P \leq 0.0002$). Further, they differed from the bLR saline group on session 1 ($P \leq 0.0001$), but did not differ from the bLR saline group on session 8. Thus, whereas dopamine may be necessary for the performance of both sign-tracking and goal-tracking CRs, it is only necessary for acquisition of a sign-tracking CR, indicating that these forms of learning are mediated by distinct neural systems.

Collectively, these data provide several lines of evidence demonstrating that dopamine does not act as a universal teaching signal in stimulus–reward learning, but selectively participates in a form of stimulus–reward learning whereby Pavlovian incentive value is attributed to a CS. First, US-evoked dopamine release in the nucleus accumbens decreased during training in sign-trackers, but not in goal-trackers. Thus, during the acquisition of a goal-tracking CR, there is not a dopamine-mediated prediction-error teaching signal because, by definition, prediction errors become smaller as delivered rewards become better predicted. Second, the CS evoked dopamine release in both sign- and goal-tracking rats, but this signal increased to a greater degree in sign-trackers, which attributed incentive salience to the CS. These data indicate that the strength of the CS–US association is reflected by dopamine release to the CS only in some forms of stimulus–reward learning. Third, bHR rats that underwent Pavlovian training in the presence of a dopamine receptor antagonist did not acquire a sign-tracking CR, consistent with previous reports[8]; however, dopamine antagonism had no effect on learning a goal-tracking CR in bLR rats. Thus, learning a goal-tracking CR does not require intact dopamine transmission, whereas learning a sign-tracking CR does.

The attribution of incentive salience is the product of previous experience (that is, learned associations) interacting with an individual's genetic propensity and neurobiological state[8,17,25–27]. The selectively bred rats used in this study have distinctive behavioural phenotypes, including greater behavioural disinhibition and reduced impulse control in bHR rats[16]. Moreover, in these lines, unlike in outbred rats[14,28], there is a strong correlation between locomotor response to novelty and the tendency to sign-track[16]. These behavioural phenotypes are accompanied by baseline differences in dopamine transmission, with bHR rats showing elevated sensitivity to dopamine agonists, increased proportion of striatal D2 receptors in a high-affinity state, greater frequency of spontaneous dopamine transients[16], and higher reward-related dopamine release before conditioning, all of which could enhance their attribution of incentive salience to reward cues[29,30]. However, basal differences in dopaminergic tone do not provide the full explanation for differences in learning styles and associated dopamine responsiveness. Outbred rats with similar baseline locomotor activity[14] and similar baseline levels of reward-related

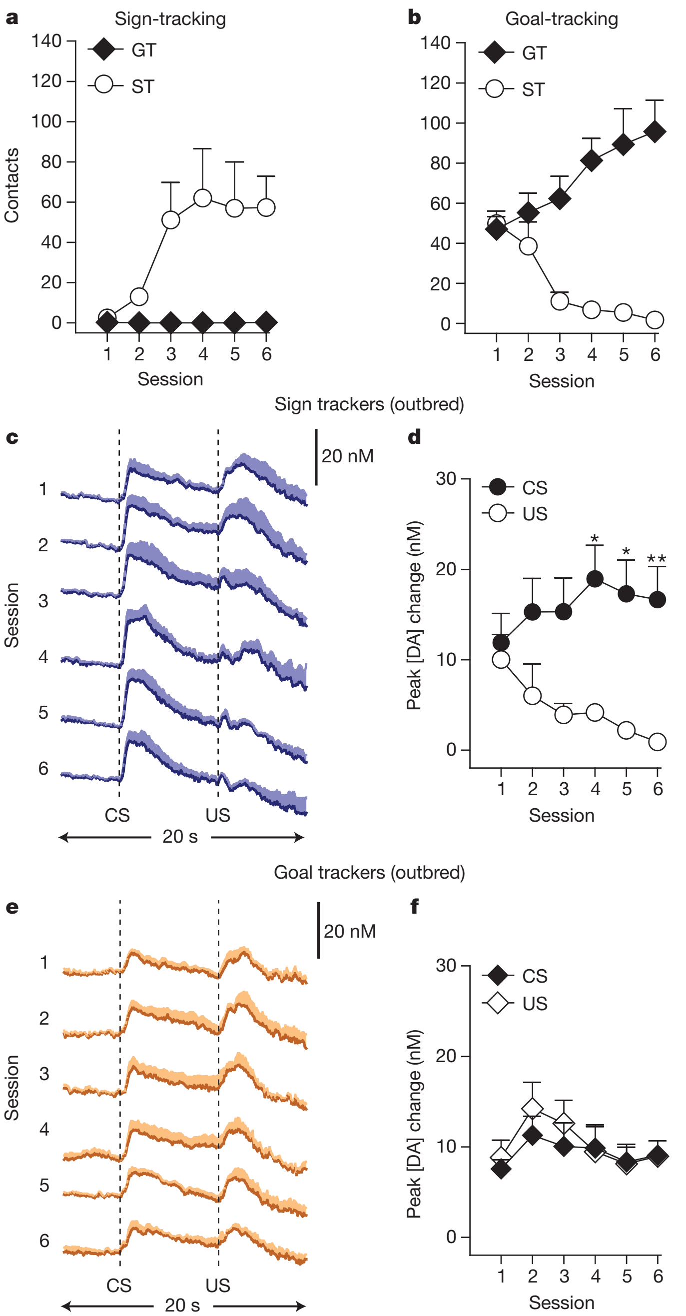


Figure 3 | Conditional responses and phasic dopamine signalling in response to CS and US presentation in outbred rats. Phasic dopamine release was recorded in the core of the nucleus accumbens using FSCV across six days of training. **a**, **b**, Behaviour directed towards the lever-CS (sign-tracking) (**a**) and behaviour directed towards the food-tray (goal-tracking) (**b**) during conditioning. Learning was evident in both groups because there was a significant effect of session both for rats that learned a sign-tracking response ($n = 6$; session effect on lever contacts: $F_{(5,25)} = 11.85$, $P = 0.0001$) and for rats that learned a goal-tracking response ($n = 5$; session effect on food-receptacle contacts: $F_{(5,20)} = 3.09$, $P = 0.03$). **c**, **e**, Change in dopamine concentration (mean + s.e.m.) in response to CS and US presentation for each session of conditioning. **d**, **f**, Change in peak amplitude (mean + s.e.m.) of the dopamine signal observed in response to CS and US presentation for each session of conditioning. (Bonferroni post-hoc comparison between CS- and US-evoked dopamine release: $*P < 0.05$; $**P < 0.01$). Panels **c** and **d** demonstrate that animals developing a sign-tracking CR ($n = 6$) show increasing phasic dopamine responses to CS presentation and decreasing responses to US presentation consistent with bHR rats. Panels **e**–**f** demonstrate that animals developing a goal-tracking CR ($n = 5$) maintain phasic responses to US presentation consistent with bLR rats.

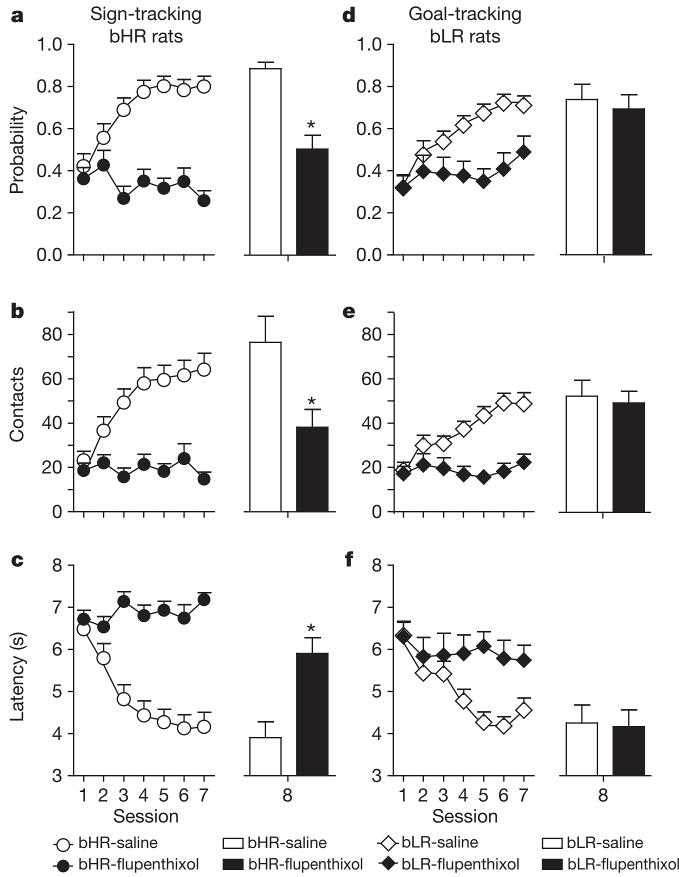


Figure 4 | Dopamine is necessary for learning CS–US associations that lead to sign-tracking, but not goal-tracking. **a**–**c**, The effects of flupenthixol on sign-tracking. **a**, Probability of approaching the lever-CS. **b**, Number of contacts with the lever-CS. **c**, Latency to contact the lever-CS. **d**–**f**, The effects of flupenthixol on goal-tracking. **d**, Probability of approaching the food-tray during lever-CS presentation. **e**, Number of contacts with the food-tray during lever-CS presentation. **f**, Latency to contact the food-tray during lever-CS presentation. Data are expressed as mean + s.e.m. Flupenthixol (sessions 1–7) blocked the performance of both sign-tracking and goal-tracking CRs. To determine whether flupenthixol influenced performance or learning of a CR, behaviour was examined following a saline injection on session 8 for all rats. bLR rats that were treated with flupenthixol before sessions 1–7 ($n = 16$) responded similarly to the bLR saline group ($n = 10$) on all measures of goal-tracking behaviour on session 8, whereas bHR rats treated with flupenthixol ($n = 22$) differed significantly from the bHR saline group ($n = 10$) on session 8 ($*P < 0.01$, saline versus flupenthixol). Thus, bLR rats learned the CS–US association that produced a goal-tracking CR even though the drug prevented the expression of this behaviour during training. Parenthetically, bHR rats treated with flupenthixol did not develop a goal-tracking CR.

dopamine release in the nucleus accumbens (see Fig. 3), differ in whether they are prone to learn a sign-tracking or goal-tracking CR, but they still develop patterns of dopamine release specific to that CR. Therefore, it appears that different mechanisms control basal dopamine neurotransmission versus the unique pattern of dopamine responsiveness to a reward cue.

The neural mechanisms underlying sign- and goal-tracking behaviour remain to be elucidated. Here we have shown that stimulus–reward associations that produce different CRs are mediated by different neural circuitry. Previous research using site-specific dopamine antagonism[21] and dopamine-specific lesions[22] indicated that dopamine acts in the nucleus accumbens core to support the learning and performance of sign-tracking behaviour. This work demonstrates that dopamine-encoded prediction-error signals are indeed present in the nucleus accumbens of sign-trackers, but not in the nucleus accumbens of goal-trackers. Although these neurochemical data alone do not rule out the possibility that prediction-error signals are present in other

dopamine terminal regions, the results from systemic dopamine antagonism demonstrate that intact dopamine transmission is generally not required for learning of a goal-tracking CR.

We thus show that dopamine is an integral part of stimulus–reward learning that is specifically associated with the attribution of incentive salience to reward cues. Individuals who attribute reward cues with incentive salience find it more difficult to resist such cues, a feature associated with reduced impulse control[16,31]. Human motivated behaviour is subject to a wide span of individual differences ranging from highly deliberative to highly impulsive actions directed towards the acquisition of rewards[32]. This work provides insight into the biological basis of these individual differences, and may provide an important step for understanding and treating impulse-control problems that are prevalent across several psychiatric disorders.

METHODS SUMMARY

The majority of these studies were conducted with adult male Sprague–Dawley rats from a selective-breeding colony which has been previously described[15]. The data presented here were obtained from bHR and bLR rats from generations S18–S22. Equipment and procedures for Pavlovian conditioning have been described in detail elsewhere[14,16]. Selectively bred rats from generations S18, S20 and S21 were transported from the University of Michigan to the University of Washington for the FSCV experiments. During each behaviour session, chronically implanted microsensors, placed in the core of the nucleus accumbens, were connected to a head-mounted voltammetric amplifier for detection of dopamine by FSCV[24]. Voltammetric scans were repeated every 100 ms to obtain a sampling rate of 10 Hz. Voltammetric analysis was carried out using software written in LabVIEW (National Instruments). On completion of the FSCV experiments, recording sites were verified using standard histological procedures. To examine the effects of flupenthixol (Sigma; dissolved in 0.9% NaCl) on the performance of sign-tracking and goal-tracking behaviour, rats received an injection (intraperitoneal, i.p.) of 150, 300 or 600 $\mu g\,kg^{-1}$ of the drug one hour before Pavlovian conditioning sessions 9, 11 and 13. Doses of the drug were counterbalanced between groups and interspersed with saline injections (i.p., 0.9% NaCl; before sessions 8, 10, 12 and 14) to prevent any cumulative drug effects. To examine the effects of flupenthixol on the acquisition of sign-tracking and goal-tracking behaviour, rats received an injection (i.p.) of either saline or 225 $\mu g\,kg^{-1}$ of the drug one hour before Pavlovian conditioning sessions 1–7.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 May; accepted 15 October 2010.
Published online 8 December 2010.

1. Schultz, W. Behavioral theories and the neurophysiology of reward. *Annu. Rev. Psychol.* **57,** 87–115 (2006).
2. Wise, R. A. Dopamine, learning and motivation. *Nature Rev. Neurosci.* **5,** 483–494 (2004).
3. Day, J. J., Roitman, M. F., Wightman, R. M. & Carelli, R. M. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nature Neurosci.* **10,** 1020–1028 (2007).
4. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275,** 1593–1599 (1997).
5. Montague, P. R., Dayan, P. & Sejnowski, T. J. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16,** 1936–1947 (1996).
6. Waelti, P., Dickinson, A. & Schultz, W. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* **412,** 43–48 (2001).
7. Balleine, B. W., Daw, N. D. & O'Doherty, J. P. in *Neuroeconomics: Decision Making and the Brain* (eds Glimcher, P. W., Camerer, C. F., Fehr, E. & Poldrack, R. A.) 367–389 (Academic Press, 2008).
8. Berridge, K. C. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology* **191,** 391–431 (2007).
9. Berridge, K. C. & Robinson, T. E. What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res. Brain Res. Rev.* **28,** 309–369 (1998).
10. Berridge, K. C., Robinson, T. E. & Aldridge, J. W. Dissecting components of reward: 'liking', 'wanting', and learning. *Curr. Opin. Pharmacol.* **9,** 65–73 (2009).
11. Panksepp, J. Affective consciousness: core emotional feelings in animals and humans. *Conscious. Cogn.* **14,** 30–80 (2005).
12. Hearst, E. & Jenkins, H. *Sign-Tracking: The Stimulus-Reinforcer Relation and Directed Action* (Monograph of the Psychonomic Society, 1974).
13. Boakes, R. in *Operant-Pavlovian Interactions* (eds Davis, H. & Hurwitz, H. M. B.) 67–97 (Erlbaum, 1977).
14. Robinson, T. E. & Flagel, S. B. Dissociating the predictive and incentive motivational properties of reward-related cues through the study of individual differences. *Biol. Psychiat.* **65,** 869–873 (2009).
15. Stead, J. D. *et al.* Selective breeding for divergence in novelty-seeking traits: heritability and enrichment in spontaneous anxiety-related behaviors. *Behav. Genet.* **36,** 697–712 (2006).
16. Flagel, S. B. *et al.* An animal model of genetic vulnerability to behavioral disinhibition and responsiveness to reward-related cues: implications for addiction. *Neuropsychopharmacology* **35,** 388–400 (2010).
17. Berridge, K. C. in *Psychology of Learning and Motivation* (ed. Medin, D. L.) 223–278 (Academic Press, 2001).
18. Cardinal, R. N., Parkinson, J. A., Hall, J. & Everitt, B. J. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.* **26,** 321–352 (2002).
19. Kelley, A. E. Functional specificity of ventral striatal compartments in appetitive behaviors. *Ann. NY Acad. Sci.* **877,** 71–90 (1999).
20. Dalley, J. W. *et al.* Time-limited modulation of appetitive Pavlovian memory by D1 and NMDA receptors in the nucleus accumbens. *Proc. Natl Acad. Sci. USA* **102,** 6189–6194 (2005).
21. Di Ciano, P., Cardinal, R. N., Cowell, R. A., Little, S. J. & Everitt, B. J. Differential involvement of NMDA, AMPA/kainate, and dopamine receptors in the nucleus accumbens core in the acquisition and performance of pavlovian approach behavior. *J. Neurosci.* **21,** 9471–9477 (2001).
22. Parkinson, J. A. *et al.* Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive Pavlovian approach behaviour: implications for mesoaccumbens dopamine function. *Behav. Brain Res.* **137,** 149–163 (2002).
23. Parkinson, J. A., Olmstead, M. C., Burns, L. H., Robbins, T. W. & Everitt, B. J. Dissociation in effects of lesions of the nucleus accumbens core and shell on appetitive pavlovian approach behavior and the potentiation of conditioned reinforcement and locomotor activity by D-amphetamine. *J. Neurosci.* **19,** 2401–2411 (1999).
24. Clark, J. J. *et al.* Chronic microsensors for longitudinal, subsecond dopamine detection in behaving animals. *Nature Methods* **7,** 126–129 (2010).
25. Robinson, T. E. & Berridge, K. C. The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Res. Brain Res. Rev.* **18,** 247–291 (1993).
26. Tindell, A. J., Smith, K. S., Berridge, K. C. & Aldridge, J. W. Dynamic computation of incentive salience: "wanting" what was never "liked". *J. Neurosci.* **29,** 12220–12228 (2009).
27. Zhang, J., Berridge, K. C., Tindell, A. J., Smith, K. S. & Aldridge, J. W. A neural computational model of incentive salience. *PLOS Comput. Biol.* **5,** e1000437 (2009).
28. Beckmann, J. S., Marusich, J. A., Gipson, C. D. & Bardo, M. T. Novelty seeking, incentive salience and acquisition of cocaine self-administration in the rat. *Behav. Brain Res.* **216,** 159–165 (2011).
29. Wyvell, C. L. & Berridge, K. C. Intra-accumbens amphetamine increases the conditioned incentive salience of sucrose reward: enhancement of reward "wanting" without enhanced "liking" or response reinforcement. *J. Neurosci.* **20,** 8122–8130 (2000).
30. Wyvell, C. L. & Berridge, K. C. Incentive sensitization by previous amphetamine exposure: increased cue-triggered "wanting" for sucrose reward. *J. Neurosci.* **21,** 7831–7840 (2001).
31. Tomie, A., Aguado, A. S., Pohorecky, L. A. & Benjamin, D. Ethanol induces impulsive-like responding in a delay-of-reward operant choice procedure: impulsivity predicts autoshaping. *Psychopharmacology* **139,** 376–382 (1998).
32. Kuo, W. J., Sjostrom, T., Chen, Y. P., Wang, Y. H. & Huang, C. Y. Intuition and deliberation: two systems for strategizing in the brain. *Science* **324,** 519–522 (2009).

Supplementary Information. is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by National Institutes of Health grants: R01-MH079292 (to P.E.M.P.), R01-DA027858 (to P.E.M.P.), T32-DA07278 (to J.J.C.), F32-DA24540 (to J.J.C.), R37-DA04294 (to T.E.R.), and 5P01-DA021633-02 (to T.E.R. and H.A.). The selective breeding colony was supported by a grant from the Office of Naval Research to H.A. (N00014-02-1-0879). We thank K. Berridge and J. Morrow for comments on earlier versions of the manuscript, and S. Ng-Evans for technical support.

Author Contributions S.B.F, J.J.C., T.E.R., P.E.M.P. and H.A. designed the experiments and wrote the manuscript. S.B.F., J.J.C., L.M., A.C., I.W. and C.A.A. conducted the experiments, S.M.C. oversaw the selective breeding colony, and S.B.F. and J.J.C. analysed the data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.E.M.P. (pemp@uw.edu) or H.A. (akil@umich.edu).

METHODS

Animals. Adult male Sprague-Dawley rats selectively bred for reactivity to a novel environment were used for the majority of these studies[15]. The data presented here were obtained from bHR and bLR rats from generations S18 to S22. The experiments followed the Guidelines for the Care and Use of Mammals in Neuroscience and Behavioural Research (National Research Council 2003) and the procedures were approved by the University Committee on the Use and Care of Animals. Unless otherwise indicated, rats were housed in pairs and kept on a 12-h light/12-h dark cycle (lights on 06:00 h) with controlled temperature and humidity and food and water were available *ad libitum*.

Voltammetry studies were conducted at the University of Washington using bHR and bLR rats from generations S18, S20 and S21as well as male Sprague-Dawley rats obtained from Charles River weighing between 300 g and 350 g upon arrival. These rats were housed individually and kept on a 12-h light/12-h dark cycle (lights on at 0700) with controlled temperature and humidity. Prior to behavioural training, food was restricted so that rats maintained 90% of their free-feeding body weight and water was available *ad libitum*. All animal procedures followed the University of Washington Institutional Animal Care and Use Committee guidelines.

Screening for selectively bred phenotypes. To confirm the selectively bred phenotypes, each generation of rats were screened for locomotor activity in novel test chambers at around 60 days of age, as previously described[15,33].

Pavlovian conditioning procedures. Equipment and procedures for Pavlovian conditioning have been described in detail elsewhere[14,16]. Briefly, standard Med Associates test chambers were equipped with a food-tray located in the middle of the front wall and a retractable lever located to the left or right of the food-tray (counterbalanced). The lever required only a 10-g force to operate, such that most contacts with the lever were detected and recorded as a 'lever press'. Operation of the pellet dispenser (Med Associates) delivered one 45-mg banana-flavoured food pellet (Bio-Serv) into the food-tray. Head entries into the food-tray were recorded each time the rat broke a photobeam located inside the receptacle.

All Pavlovian training sessions were conducted between 13:00 h and 18:00 h. Banana-flavoured food pellets were placed into the rats' home cages for 2 days before training to familiarize the animals with this food (the unconditioned stimulus, US). Two pre-training sessions were conducted that consisted of the delivery of 50 food pellets, which were randomly delivered on a variable-interval 30-s schedule (25-min session), during which it was determined whether the rats were reliably retrieving the food pellets. Following pre-training sessions, Pavlovian training sessions consisted of the presentation of the illuminated lever (conditioned stimulus, CS) in the chamber for 8 s, and then immediately upon its retraction a 45-mg food pellet (US) was delivered into the food-tray (the 'goal'). The CS was presented on a random-interval 90-s schedule and each Pavlovian training session consisted of 25 trials (or CS–US pairings). Training continued for 6–12 sessions. Rats in the 'random' groups received presentations of the CS and US, each on a variable-interval 90-s schedule.

The following events were recorded using Med Associates software: (1) the number of lever-CS contacts, (2) the latency to the first lever-CS contact, (3) the number of food-tray entries during lever-CS presentation, and (4) the latency to the first food-tray entry during lever-CS presentation. It is important to note that no response is required for the rat to receive the reward (US), yet distinct CRs emerge as a result of Pavlovian conditioning. The outcome measures listed above allow us to examine CS-directed (sign-tracking) versus goal-directed (goal-tracking) responses. Using these measures we calculated the probability that a rat would approach the lever-CS or the food-tray as well as the difference in its probability of approaching the lever-CS versus the food-tray.

Statistical analysis of Pavlovian conditional responses. Differences in the conditional response that emerged across training sessions were analysed using linear mixed effects models (SPSS 17.0; see also ref. 34), in which phenotype and session were treated as independent variables. In addition, the effect of session for each phenotype was analysed separately. For all analyses, the covariance structure was explored and modelled appropriately. When significant main effects or interactions were detected, Bonferroni post-hoc comparisons were made. The differences in the probability of approaching the lever-CS versus the food-tray (Fig. 1e) were further examined using one-sample *t*-tests (with hypothesized value of 0) to determine whether either phenotype exhibited a preference for the lever-CS or the food-tray.

Conditioned reinforcement test. The conditioned reinforcement test occurred one day after the last of 12 Pavlovian training sessions. The conditioned reinforcement test was conducted in the same standard Med Associates chambers as described above. However, for the purposes of this test the chambers were rearranged such that the retractable lever was placed in the centre of the front wall in between two nosepoke ports. The 'active' port was placed on the side of the wall opposite to the location of the lever-CS during Pavlovian training. During the 40-min conditioned reinforcement test nosepokes into the port designated 'active' resulted in the 2-s presentation of the illuminated lever, whereas pokes into the other 'inactive' port were without consequence. The number of nosepokes into the active and inactive ports and the number of contacts with the lever were recorded throughout the test session.

Statistical analysis of conditioned reinforcement. Performance on the conditioned reinforcement test was analysed using a three-way analysis of variance (ANOVA) in which phenotype, group (paired versus unpaired) and port (active versus inactive) were treated as independent variables and the number of pokes as the dependent variable. Further analyses were then conducted to determine the effect of group or port for each phenotype and the effect of phenotype or group for each port.

FSCV. The following procedures were in accordance with the University of Washington Institutional Animal Care and Use Committee guidelines. Surgical preparation for *in vivo* voltammetry used an aseptic technique. Rats were anaesthetized with isofluorane and placed in a stereotaxic frame. The scalp was swabbed with 10% povidone iodine, bathed with a mixture of lidocaine (0.5 mg kg^{-1}) and bupivicaine (0.5 mg kg^{-1}), and incised to expose the cranium. Holes were drilled and cleared of dura mater above the nucleus accumbens core (1.3-mm lateral and 1.3-mm rostral from the bregma), and at convenient locations for a reference electrode and three anchor screws. The reference electrode and anchor screws were positioned and secured with cranioplastic cement, leaving the working electrode holes exposed. Once the cement cured, the microsensors were attached to the voltammetric amplifier and lowered into the target recording regions (the core of the nucleus accumbens, 7.0-mm ventral of dura mater). Finally, cranioplastic cement was applied to the part of the cranium still exposed to secure the working electrode.

Voltammetric measurement. During all experimental sessions, chronically implanted microsensors were connected to a head-mounted voltammetric amplifer for dopamine detection by FSCV[24]. Voltammetric scans were repeated every 100 ms to obtain a sampling rate of 10 Hz. When dopamine is present at the surface of the electrode during a voltammetric scan, it is oxidized during the anodic sweep to form dopamine-*o*-quinone (peak reaction at approximately +0.7 V), which is reduced back to dopamine in the cathodic sweep (peak reaction at approximately −0.3 V). The ensuing flux of electrons is measured as current and is directly proportional to the number of molecules that undergo the electrolysis. The redox current obtained from each scan provides a chemical signature that is characteristic of the analyte, allowing resolution of dopamine from other substances. For quantification of changes in dopamine concentration over time, the current at its peak oxidation potential can be plotted for successive voltammetric scans. Waveform generation, data acquisition and analysis were carried out on a PC-based system using two PCI multifunction data acquisition cards and software written in LabVIEW (National Instruments).

Statistical analysis of voltammetry data. Voltammetric data analysis was carried out using software written in LabVIEW (National Instruments) and low-pass filtered at 2,000 Hz. Dopamine was isolated from the voltammetric signal with chemometric analysis[35] using a standard training set based upon stimulated dopamine release detected by chronically implanted electrodes. Dopamine concentration was estimated on the basis of the average post-implantation sensitivity of electrodes[24]. Before the generation of surface plots and analysis of peak values, all data were smoothed with a 5-point within-trial running average. Peak dopamine values in response to the US and CS were obtained by taking the largest value in the 3-s period after stimulus presentation. Peak values were then compared using mixed models ANOVA with training session as the repeated measure and stimulus (CS and US) or phenotype (bHR and bLR) as the between-group measure. Peak CS-evoked dopamine signalling was also analysed across trials using linear regression. The slopes obtained for the regression were compared between groups using independent, two-sample *t*-tests. All post-hoc comparisons were made with the Bonferroni correction for multiple tests. All statistical analyses were carried out using Prism (GraphPad Software). Voltammetric data for dopamine responses to the CS and US were also analysed using an area-under-the-curve approach. This approach did not alter the statistical effects of any comparison reported in the paper for peak dopamine value (specific statistical results not shown).

Histological verification of recording site. On completion of experimentation, animals were anesthetized with intraperitoneal ketamine (100 mg kg^{-1}) and xylazine (20 mg kg^{-1}) and then transcardially perfused with saline followed by 4% paraformaldehyde. Brains were removed and post-fixed in paraformaldehyde for 24 h and then rapidly frozen in an isopentane bath (~5 min), sliced on a cryostat (50-μm coronal sections, 20 °C) and stained with cresyl violet to aid in visualization of anatomical structures.

Effects of flupenthixol on sign-tracking and goal-tracking performance. The effects of flupenthixol (a D1/D2 antagonist; Sigma) on the performance of sign-tracking and goal-tracking behaviour were examined after seven sessions of Pavlovian conditioning. All rats received an injection (i.p.) of 150, 300 or

600 μg kg^{-1} of the drug one hour before Pavlovian conditioning sessions 9, 11 and 13. Doses of the drug (dissolved in 0.9% NaCl) were counterbalanced between groups and interspersed with saline injections (i.p., 0.9% NaCl; before sessions 8, 10, 12 and 14) to prevent any cumulative drug effects. The following measures were recorded to examine the effects of the drug on the CR: (1) the number of lever-CS contacts, (2) the latency to the first lever-CS contact, (3) the number of food-tray entries during lever-CS presentation, and (4) the latency to the first food-tray entry during lever-CS presentation. In addition, a nosepoke port was added to the test chamber on the wall opposite the retractable lever and responses into this port were recorded as an index of nonspecific activity. For all measures the response to saline was averaged (across sessions 8, 10, 12 and 14) and compared to the response following each of the three doses of flupenthixol.

Statistical analysis of effects of flupenthixol on performance of the CRs. The effects of flupenthixol on the performance of sign-tracking and goal-tracking behaviour (Supplementary Fig. 6) were analysed using linear mixed effects models with phenotype and dose treated as independent variables. Each phenotype was also analysed separately to determine the effect of dose on a given behaviour and Bonferroni post-hoc comparisons were made to determine whether behaviour at a given dose was significantly different from that in response to saline.

Effects of flupenthixol on the learning of sign-tracking and goal-tracking. The effects of flupenthixol on the acquisition of sign-tracking and goal-tracking CRs were examined using two generations of bred rats (S21 and S22). Rats received an injection of either saline (i.p.; 0.9% NaCl) or 225 μg kg^{-1} of flupenthixol one hour before Pavlovian conditioning sessions 1–7. This dose of drug was chosen based on the ‘performance’ study described above, because we wanted to avoid any nonspecific inhibitory effects on motor activity. Rats from both generations that received flupenthixol before sessions 1–7 then received an injection of saline before session 8. However, only rats from the S22 generation that received saline before sessions 1–7 also received saline before session 8. Thus, the number of rats that received saline during training and were also pretreated with saline before session 8 is lower than that for the other groups (that is, on session 8, bHR saline, $n = 10$; bLR saline, $n = 10$). The following measures were recorded and analysed to examine the effects of flupenthixol on sign-tracking and goal-tracking behaviour: (1) the number of lever-CS contacts, (2) the latency to the first lever-CS contact, (3) the number of food-tray entries during lever-CS presentation, and (4) the latency to the first food-tray entry during lever-CS presentation.

Statistical analysis of effects of flupenthixol on the learning of the CRs. Linear mixed effects models were used to examine the effects of flupenthixol on the performance and learning of sign-tracking or goal-tracking behaviour (Supplementary Fig. 6). For these analyses each phenotype was analysed separately to determine the effect of dose on a given behaviour and treatment (saline versus flupenthixol) and session (1–7) were treated as independent variables. To determine whether flupenthixol prevented the expression of the conditioned response or the learning of a conditioned response we also examined behaviour following a saline injection on session 8 (drug-free test session). Behaviour on session 8 was compared between treatment groups using an unpaired t-test for each phenotype separately. We also compared the response on session 8 of the groups that received flupenthixol during training to that of the group that received flupenthixol on session 1 (using a paired t-test) and to that of the saline control group on session 1 (using an unpaired t-test).

33. Clinton, S. M, *et al.* Individual differences in novelty-seeking and emotional reactivity correlate with variation in maternal behavior. *Horm. Behav.* **51,** 655–664 (2007).
34. Verbeke, G. & Molenberghs, G. *Linear Mixed Models for Longitudinal Data* (Springer, 2000).
35. Heien, M. L. Johnson, M. A. & Wightman, R. M. Resolving neurotransmitters detected by fast-scan cyclic voltammetry. *Anal. Chem.* **76,** 5697–5704 (2004).

doi:10.1038/nature09592

Experimental niche evolution alters the strength of the diversity–productivity relationship

Dominique Gravel[1,2], Thomas Bell[3], Claire Barbera[2], Thierry Bouvier[4], Thomas Pommier[4,5], Patrick Venail[2]† & Nicolas Mouquet[2]

The relationship between biodiversity and ecosystem functioning (BEF) has become a cornerstone of community and ecosystem ecology[1–3] and an essential criterion for making decisions in conservation biology and policy planning[4,5]. It has recently been proposed that evolutionary history should influence the BEF relationship because it determines species traits and, thus, species' ability to exploit resources[6,7]. Here we test this hypothesis by combining experimental evolution with a BEF experiment. We isolated 20 bacterial strains from a marine environment and evolved each to be generalists or specialists[8]. We then tested the effect of evolutionary history on the strength of the BEF relationship with assemblages of 1 to 20 species constructed from the specialists, generalists and ancestors[9]. Assemblages of generalists were more productive on average because of their superior ability to exploit the environmental heterogeneity[10]. The slope of the BEF relationship was, however, stronger for the specialist assemblages because of enhanced niche complementarity. These results show how the BEF relationship depends critically on the legacy of past evolutionary events.

Two fundamental ecological mechanisms can generate positive BEF relationships[11,12]. First, species may occupy complementary ecological niches, for example by feeding on different resources. In communities of complementary species, more of the total available niche space is filled in diverse communities, resulting in better community-wide resource use. Second, high-functioning and competitively dominant species are more likely to be found within species-rich communities (the sampling effect). Both mechanisms require a detailed understanding of species' phenotypic traits[13,14]. There has been, however, virtually no effort to understand how the evolution of species traits within ecological communities affects ecosystem functioning[6,7,15].

One important trait that determines complementarity is the degree of resource specialization, that is, the number of resources a species is able to exploit. Species niche width will tend to evolve to match the amount of available environmental variation[16,17]. In simple environments specialized types are expected to evolve, whereas generalists are more likely to appear in environments containing many resources[18,19]. The degree of specialization could alter the BEF relationship[10], so a full understanding the relationship must account for the evolutionary forces driving trait diversity.

In communities containing only specialist species that feed on different resources, the species do not compete with each other and their effects on ecosystem functioning (here productivity) are therefore additive (that is, the BEF relationship is linear; Fig. 1). However, the increased ability to exploit one resource might come with a lower ability to exploit any other, that is, there is a trade-off between resource usage ability[18] (Fig. 1a). For any type of trade-off, evolution towards generalization will affect the BEF relationship (Fig. 1b and Supplementary Information, section 1). First, ecosystem functioning at low diversity should be lower for specialists because they are inefficient at exploiting environmental heterogeneity. Second, generalization should reduce the contribution of additional species to ecosystem functioning and, thus, the slope (that is, strength) of the BEF relationship. Generalization also increases niche overlap and thus produces a nonlinear, saturating BEF relationship. We therefore predicted that the ecosystem functioning would be higher for generalists at low diversity and that the slope of the BEF relationship would be reduced in communities of generalists.

We tested these predictions by experimentally evolving niche specialization and conducting BEF experiments. Briefly, we promoted the evolution of generalist and specialist strategies from 20 ancestral bacterial strains that had been isolated from a marine environment. Each strain was grown either on a single resource (one different carbon substrate for each strain) or on a mixture of 31 resources (with a total resource availability equal to that in the single-resource treatment). Bacteria were serially transferred to fresh medium every 48 h for 32 transfers, allowing evolutionary adaptations[20] (Supplementary Information, section 2). Bacteria grown on the mixed medium (hereafter called generalists) tended to have higher performance on a wide array of substrates in comparison with the bacteria evolved on the simple medium consisting of a single resource (hereafter called specialists; see below).

We conducted the BEF experiment on the mixed medium of 31 resources for both evolutionary schemes (generalist and specialist) and with the ancestral strains (hereafter referred as species). We used the productivity (bacterial metabolic activity) after 48 and 72 h as our measure of ecosystem functioning[15]. The selection period resulted in a substantially increased productivity at all levels of species richness for both specialists and generalist assemblages (Fig. 2). Productivity significantly increased with the logarithm of species richness (48 h, $F_{1,1,506} = 291.2$, $P < 0.001$; 72 h, $F_{1,1,506} = 179.2$, $P < 0.001$), indicating that productivity was a saturating function of species richness. Productivity in monocultures differed significantly among treatments (48 h, $F_{2,1,506} = 1{,}751.9$, $P < 0.001$; 72 h, $F_{2,1,506} = 2{,}309.3$, $P < 0.001$), with ancestors performing the worst, followed in order by specialists and generalists. The slope of the BEF relationship also differed significantly among treatments (48 h, $F_{2,1,506} = 16.2$, $P < 0.001$; 72 h, $F_{1,1,506} = 8.8$, $P < 0.001$), being steeper for specialists. A model accounting for species composition also showed that the BEF relationship of ancestors was best described by a linear function of species richness and a nonlinear (saturating) function for the two evolutionary treatments (Supplementary Information, section 3). There was no relationship between the contribution of ancestors to the BEF and the contribution of their evolved counterparts (Supplementary Information, section 4). The difference between the slopes of the specialist and generalist BEF relationships was even stronger after 72 h, a result similar to those of experiments conducted with plant[21,22] and marine[23] communities.

We investigated whether the difference in the strength of the BEF relationship resulted from specialization, by growing each of the ancestral strains and specialist and generalist lineages on the 31 individual carbon substrates to estimate their final niche width. We recorded the number

[1]Université du Québec à Rimouski, Département de Biologie, Chimie et Géographie, 300 Allée des Ursulines, Québec G5L 3A1, Canada. [2]Institut des Sciences de l'Evolution UMR 5554, Centre National de la Recherche Scientifique, Université Montpellier 2, CC 065, Place Eugène Bataillon, 34095 Montpellier Cedex 05, France. [3]Department of Zoology, University of Oxford, Oxford OX1 3PS, UK. [4]Ecosystèmes Lagunaires UMR 5119, Centre National de la Recherche Scientifique, Université Montpellier 2, CC 093, Place Eugène Bataillon, 34095 Montpellier Cedex 05, France. [5]Laboratoire d'Ecologie Microbienne (UMR 5557, USC 1193), Université Lyon I, INRA, CNRS, Bâtiment G. Mendel, 43 Boulevard du 11 Novembre 1918, 69622 Villeurbanne, France. †Present address: Centro de Investigaciones Microbiológicas, Universidad de los Andes, Carrera 1A No. 18A-10, Oficina A305, Bogotá, Colombia.

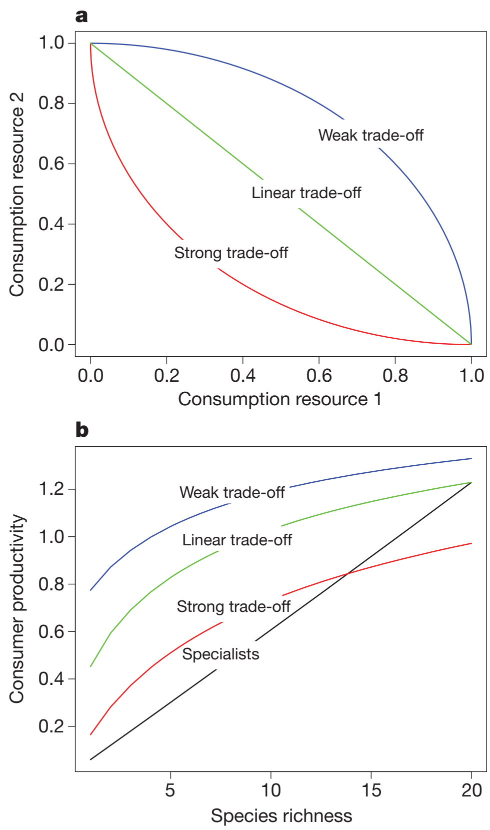


Figure 1 | Theoretical predictions of the effect of niche specialization on the strength of the BEF relationship. **a**, Hypothetical relationship between the consumption rates of a generalist on two resources. A perfect trade-off follows a straight line and any deviation to this reflects a relative cost (strong trade-off) or benefit (weak trade-off) of generalization. **b**, The BEF relationship will be affected by generalization and the type of trade-off. The figure represents a simulation experiment with $N = 20$ resources (R_i) and up to as many ($S \leq N$) consumer species (C_j). The dynamics of this system is given by the simple chemostat model $\mathrm{d}R_i/\mathrm{d}t = eR_0 - eR_i - \sum_{j=1}^{S} \alpha_{ij} R_i C_j$, $\mathrm{d}C_j/\mathrm{d}t = \sum_{i=1}^{N} \alpha_{ij} R_i C_j - mC_j - eC_j$, where e is the dilution rate, R_0 is the resource concentration in the inflow (for simplicity, we suppose that all resources have the same concentration in the inflow), α_{ij} is the per capita consumption rate of the resource i by consumer j, and m is the mortality rate. The ecosystem productivity at equilibrium is equivalent to the total resource consumption and is given by $\Phi = \sum_i^N \sum_j^S \alpha_{ij} R_i C_j$. Our simulation parameters are $R_0 = 1$, $e = 0.1$ and $m = 0.1$. We specified for the specialist that the consumption rate for its preferential resource is $\alpha_{ii} = 1$ and $\alpha_{i,j \neq i} = 0$ for the alternative resources. For the generalists, we specified the performance on the preferential resource to be $\alpha_{ii} = 0.6$ and an equal partitioning of the consumption rates between alternative resources that sums to 0.4 for the linear trade-off, 0.1 for the strong trade-off and 0.55 for the weak trade-off. We simulated communities of 1 to 20 species for specialists and generalists with linear, weak and strong trade-offs. Further details of the model and analytical results are given in Supplementary Information.

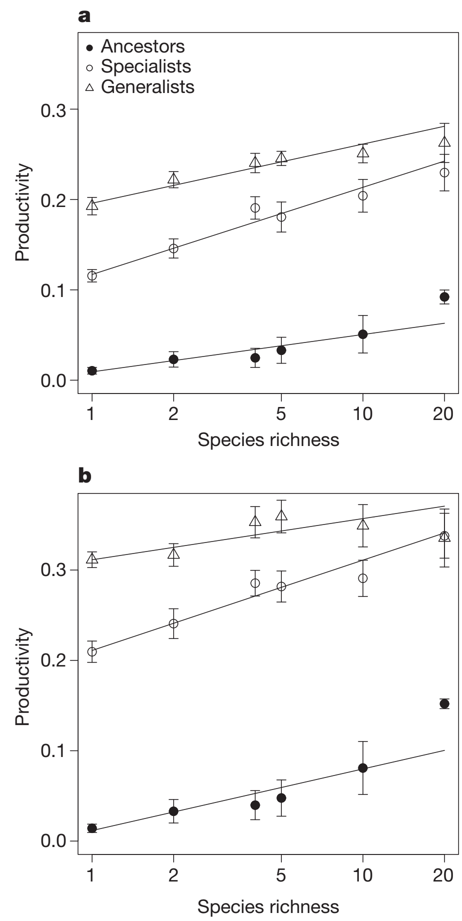


Figure 2 | Evolutionary treatments affect the strength of the biodiversity–productivity relationships. Productivity is measured as the absorbance at 590 nm after 48 h (corresponding to the conditions experienced during the selection experiment; **a**) and 72 h (**b**). Data show mean ± s.e. ($n = 504$ per evolutionary treatment). Lines depict the results of the analysis of covariance model.

of substrates each lineage was able to exploit. Bacteria cultured on single substrates adapted to fewer substrates than bacteria cultured on mixed medium (Table 1 and Supplementary Information, section 5). The generalists were able on average to exploit a larger number of substrates (10.75 ± 1.49 (s.e.) of the 31 substrates) than the specialists (4.80 ± 0.51 (s.e.); t-test for paired samples, $t_{19} = 3.45$, $P = 0.002$). The average number of shared substrates between all pairs of strains and lineages was also much higher for generalists than for ancestors and specialists (Table 1). This ability to exploit more substrates is reflected in the performance of the bacteria when grown on the mixed medium. The maximal performance recorded for each substrate for generalists was also significantly higher than the maximal performance for specialists (t-test for paired samples, $t_{30} = 2.95$, $P = 0.006$; Table 1), which suggests there was no trade-off in resource usage ability. It has been shown that concurrent adaptation to multiple resources does not always limit the capability to exploit each resource individually[8]. In fact, mutations increasing fitness on a given resource can even sometimes increase fitness on other resources (that is, synclinal selection[24]), preventing the occurrence of trade-offs. We also checked for within-lineage genotypic variability and found it to be low (Supplementary Information, section 6). Most generalist lineages were composed of generalist genotypes, but some consisted of mixtures of coexisting specialist genotypes. However, the qualitative results and conclusions (Fig. 2) were unaffected when the analysis was restricted to only those lineages that were composed of the most generalist genotypes (Supplementary Information, section 6).

Some lineages that evolved on the mixed medium responded weakly to the experimental evolution and remained specialists, and some lineages from the specialist treatment evolved towards generalization (Supplementary Information, section 5). Consequently, there is variance in ecosystem functioning that should be better predicted by functional diversity[7]. In addition to enhanced complementarity, better performance of the generalists on each carbon source might also have contributed to the difference in the BEF relationships between evolutionary treatments (Fig. 2). We calculated a niche diversity index (NDI) for every assemblage of the specialist and generalist treatments. The NDI is the number of different substrates a community is able to exploit, calculated on the basis of assays of individual species[25] (Supplementary Information, section 7). Given that generalists were more productive on each substrate, we expected that for a given NDI

Table 1 | The effect of experimental evolution on specialization

	Ancestral strains	Specialist lineages	Generalist lineages
No. of substrates used	4.00 ± 0.69	4.80 ± 0.51	10.75 ± 1.49
No. of substrates used by the 20 species assemblages	16	14	29
Average no. of shared substrates between strain–lineage pairs	2.25 ± 0.08	2.74 ± 0.11	5.80 ± 0.24
Average maximal productivity for each substrate	0.21 ± 0.07	0.23 ± 0.08	0.42 ± 0.10
Productivity of monocultures on the mixed medium	0.01 ± 0.01	0.12 ± 0.01	0.19 ± 0.02

Niche specialization was assessed from performance assays on the 31 carbon sources. A strain (or lineage) was considered able to exploit a substrate when its absorbance at 590 nm after 48 h was larger than the 95% distribution of the blanks. All data, mean ± s.e.

productivity would be larger for the generalist assemblages. We found that productivity significantly increases with community NDI ($F_{1,667} = 84.0$, $P < 0.001$; Fig. 3) (we note that the analysis is conducted on the NDI range from 3 to 14 to meet the requirements of analysis of covariance) and the logarithm of species richness ($F_{1,667} = 5.23$, $P = 0.022$). The intercepts and the slope of the NDI–productivity relationship differ between evolutionary treatments (48 h, $F_{1,667} = 72.6$, $P < 0.001$; 72 h, $F_{1,667} = 19.4$, $P < 0.001$). Overall, most of the variance is accounted for by the NDI and the effect of the evolutionary treatments on the intercept. The larger amount of explained variance by the NDI argues for complementarity as the dominant mechanism. Our experimental evolutionary treatments therefore affected both species complementarity and maximal productivity at equivalent complementarity.

The ancestral strains had not previously encountered the experimental conditions, so it is unsurprising that the intercept of the BEF relationship was greatly reduced as a result of their maladaptation (Fig. 2). Nonetheless, there is still a significant positive BEF relationship. Because each species was equally represented in the experiment, it is possible to estimate the degree to which they were associated with higher- or lower-than-average levels of functioning[26]. We found that the inferred species contributions were dominated by a single ancestral strain, whereas they were distributed more equitably in the specialist and the generalist treatments (Supplementary Information, sections 3 and 4). We note there is no significant correlation between the ancestors' contribution to the BEF and the contribution of their evolved counterparts. The data therefore provide evidence that evolutionary history could affect both mechanisms of the BEF relationship. First, the BEF relationship will be stronger for communities of specialists because of enhanced complementarity. Second, if most species are maladapted, few species are able to contribute to functioning and sampling effects dominate. Such a mechanism might be particularly important for ecosystem functioning in variable environments, where species are far from their optimal fitness peaks[27].

In this study, we have deliberately evolved independent lineages of specialists and generalists to compare assemblages of species that come from the same ancestral strain but have different evolutionary histories. It is likely that in nature species will diversify in complex assemblages of specialists and generalists depending on the environmental context[28]. We note that we cannot exclude the possibility that the generalists were an ensemble of specialist genotypes. Previous work has, however, shown that selection in a heterogeneous environment is most likely to result in the evolution of generalists[8]. In any case, the genotypic variability in a population, leading to increased generality, is still a species' trait that will influence the BEF relationship. Some of the changes observed between the evolved lineages might also have come through physiological adaptation. However, after at least a hundred generations we have found highly contrasting metabolic profiles (Table 1), no correlation between species contributions to the BEF relationship (Supplementary Information, section 3) and a strong response to selection (Supplementary Information, section 2). All of these observations are consistent with evolutionary changes.

A variety of BEF relationships have been observed and different ecological mechanisms have been inferred[22,29,30]. Our results provide strong support for the role of complementarity and evolutionary history in BEF. We found that specialists contribute more to the BEF. Monocultures of generalists were also found among the most productive assemblages. For conservation decisions, these results emphasize that, on average, the loss of specialists will have stronger effects on ecosystem functioning, but that losing a generalist species might have disproportionate effects when there is low redundancy. Our understanding of the mechanisms underlying the BEF relationship has now moved to a point where we cannot only distinguish among mechanisms, but can also manipulate these mechanisms experimentally. Investigations should now turn to understanding the evolutionary pressures that maintain niche diversification in natural communities, along with the trade-offs involved, and their effect on the BEF relationship.

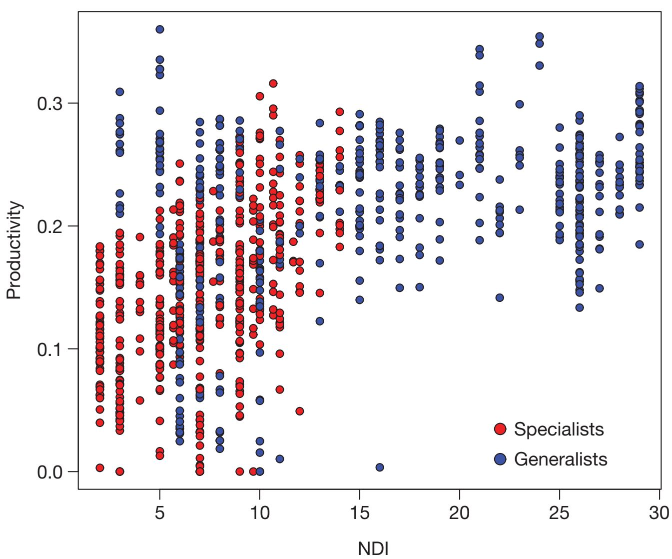


Figure 3 | The relationship between NDI and ecosystem functioning. NDI is the total number of carbon substrates a community is able to exploit, assessed from the individual ability of each lineage to exploit the carbon substrates and the community composition.

METHODS SUMMARY

We isolated 31 phage-free bacterial strains from coastal sea water sampled off the Bay of Blanes, Spain, on the basis of their morphologies. We sequenced the 16S ribosomal DNA genes of the ancestral strains to confirm that different taxa were used in the experiment (Supplementary Information, section 8). Each strain individually underwent selection on a different, single-carbon substrate of an EcoPlate to obtain specialists, and underwent selection on a highly mixed medium made from a mixture of all 31 EcoPlate carbon substrates to obtain generalists. We transferred the bacteria to a fresh medium every 48 h during 64 d of incubation at 20 °C. After the selection period, we kept 20 lineages from among those that persisted and conducted a BEF experiment. We assembled microcosms with diversity levels, s, of 1, 2, 4, 5, 10 and 20 species for each of the three treatments (ancestral strains and specialist and generalist lineages). For each evolutionary treatment and diversity level, we created $20/s$ different assemblages by randomly selecting species from the species pool without replacement (for example, if $s = 5$ we randomly assigned the 20 species to four assemblages)[9], for a total of 42 assemblages. We carried out this process independently four times, so there were a total of 168 different assemblages for each evolutionary treatment. Each assemblage was replicated three times for a total of 1,512 microcosms (the product of three treatments, three replicates and 168 assemblages). We measured the light absorbance at 590 nm after 48 and 72 h to approximate productivity (reported

values were corrected by removing the average value of the blanks). We also conducted assays for each ancestral strain and the specialist and generalists lineages by incubating them on the 31 carbon substrates for 48 h at 20 °C. The assays were used to quantify generality and the niche diversity in the assemblages.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 September; accepted 20 October.

Published online 5 December 2010.

1. Kinzig, A. P., Pacala, S. & Tilman, G. D. (eds) *The Functional Consequences of Biodiversity: Empirical Progress and Theoretical Extensions* (Princeton Univ. Press, 2002).
2. Loreau, M., Naeem, S. & Inchausti, P. (eds) *Biodiversity and Ecosystem Functioning: Synthesis and Perspectives* (Oxford Univ. Press, 2002).
3. Naeem, S., *et al. Biodiversity, Ecosystem Functioning, and Human Wellbeing: An Ecological and Economic Perspective* (Oxford Univ. Press, 2009).
4. Srivastava, D. S. & Vellend, M. Biodiversity-ecosystem function research: is it relevant to conservation? *Annu. Rev. Ecol. Evol. Syst.* **36,** 267–294 (2005).
5. Duffy, J. E. Why biodiversity is important to the functioning of real-world ecosystems. *Front. Ecol. Environ.* **7,** 437–444 (2009).
6. Cadotte, M. W., Cardinale, B. J. & Oakley, T. H. Evolutionary history and the effect of biodiversity on plant productivity. *Proc. Natl Acad. Sci. USA* **105,** 17012–17017 (2008).
7. Cadotte, M. W., Cavender-Bares, J., Tilman, D. & Oakley, T. H. Using phylogenetic, functional and trait diversity to understand patterns of plant community productivity. *PLoS ONE* **4,** e5695 (2009).
8. Barrett, R. D. H., MacLean, R. C. & Bell, G. Experimental evolution of *Pseudomonas fluorescens* in simple and complex environments. *Am. Nat.* **166,** 470–480 (2005).
9. Bell, T. *et al.* The contribution of species richness and composition to bacterial services. *Nature* **436,** 1157–1160 (2005).
10. Loreau, M. Microbial diversity, producer-decomposer interactions and ecosystem processes: a theoretical model. *Proc. R. Soc. Lond. B* **268,** 303–309 (2001).
11. Tilman, D., Lehman, C. L. & Thomson, K. T. Plant diversity and ecosystem productivity: theoretical considerations. *Proc. Natl Acad. Sci. USA* **94,** 1857–1861 (1997).
12. Loreau, M. Biodiversity and ecosystem functioning: a mechanistic model. *Proc. Natl Acad. Sci. USA* **95,** 5632–5636 (1998).
13. Diaz, S. & Cabido, M. Vive la difference: plant functional diversity matters to ecosystem processes. *Trends Ecol. Evol.* **16,** 646–655 (2001).
14. Reiss, J., Bridle, J. R., Montoya, J. M. & Woodward, G. Emerging horizons in biodiversity and ecosystem functioning research. *Trends Ecol. Evol.* **24,** 505–514 (2009).
15. Venail, P. A. *et al.* Diversity and productivity peak at intermediate dispersal rate in evolving metacommunities. *Nature* **452,** 210–214 (2008).
16. Levins, R. *Evolution in Changing Environments: Some Theoretical Considerations* (Princeton Univ. Press, 1968).
17. Futuyma, D. J. & Moreno, G. The evolution of ecological specialization. *Annu. Rev. Ecol. Syst.* **19,** 207–233 (1988).
18. Kassen, R. The experimental evolution of specialists, generalists, and the maintenance of diversity. *J. Evol. Biol.* **15,** 173–190 (2002).
19. MacLean, R. C. Adaptive radiation in microbial microcosms. *J. Evol. Biol.* **18,** 1376–1386 (2005).
20. Buckling, A., Maclean, R. C., Brockhurst, M. A. & Colegrave, N. The Beagle in a bottle. *Nature* **457,** 824–829 (2009).
21. Tilman, D. *et al.* Diversity and productivity in a long-term grassland experiment. *Science* **294,** 843–845 (2001).
22. Cardinale, B. J. *et al.* Impacts of plant diversity on biomass production increase through time because of species complementarity. *Proc. Natl Acad. Sci. USA* **104,** 18123–18128 (2007).
23. Stachowicz, J. J., Graham, M., Bracken, M. E. S. & Szoboszlai, A. I. Diversity enhances cover and stability of seaweed assemblages: the role of heterogeneity and time. *Ecology* **89,** 3008–3019 (2008).
24. MacLean, R. C., Bell, G. & Rainey, P. B. The evolution of a pleiotropic fitness tradeoff in *Pseudomonas fluorescens*. *Proc. Natl Acad. Sci. USA* **101,** 8072–8077 (2004).
25. Salles, J. F., Poly, F., Schmid, B. & Le Roux, X. Community niche predicts the functioning of denitrifying bacterial assemblages. *Ecology* **90,** 3324–3332 (2009).
26. Bell, T. *et al.* A linear model method for biodiversity-ecosystem functioning experiments. *Am. Nat.* **174,** 836–849 (2009).
27. Yachi, S. & Loreau, M. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proc. Natl Acad. Sci. USA* **96,** 1463–1468 (1999).
28. Abrams, P. A. Adaptive change in the resource-exploitation traits of a generalist consumer: the evolution and coexistence of generalists and specialists. *Evolution* **60,** 427–439 (2006).
29. Balvanera, P. *et al.* Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecol. Lett.* **9,** 1146–1156 (2006).
30. Bell, T. *et al.* in *Biodiversity, Ecosystem Functioning and Human Well-being* (eds Naeem, S. *et al.*) 121–133 (Oxford Univ. Press, 2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Barraclough, G. Bell, M. Loreau, R. C. MacLean and A. Paquette for comments on earlier versions of the manuscript. This work was supported by a fellowship from the Natural Sciences and Engineering Research Council of Canada and a research grant from the Canada Research Chair Program, to D.G., and research grant ANR-09-JCJC-0110-01, to N.M.

Author Contributions D.G., T. Bell, C.B., T. Bouvier, T.P., P.V. and N.M. designed the research; D.G., C.B., T.P., P.V., T. Bell and N.M. conducted the research; and D.G., T. Bell and N.M. contributed to the model, analytical tools and wrote the manuscript. D.G., T. Bell, C.B., T. Bouvier, T.P., P.V. and N.M. edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.G. (dominique_gravel@uqar.qc.ca) or N.M. (nmouquet@univ-montp2.fr).

METHODS

Bacterial isolation. Bacterial strains were isolated from coastal sea water sampled from the Bay of Blanes, Spain (40° 40′ N, 2° 50′ E) on 20–21 September 2007. Five replicate samples of 100 μl of sea water were spread on marine agar plates (BD Difco Marine Agar 2216; autoclaved for 20 min at 121 °C) and grown for 5 d at 12 °C, which was the *in situ* temperature at sampling time. Ninety-five colonies with distinct morphotypes (that is, size, shape and colour) were isolated over four weeks, clean-streaked three times and frozen in glycerol at −80 °C.

We then sequenced the 16S rDNA genes of the ancestral strains to confirm different taxa. A single colony of the 95 isolates was picked and dissolved in 10 μl TE (10 nM Tris, 1mM EDTA, pH 8.0) buffer, heated for 5 min at 95 °C and centrifuged briefly. The supernatant (1 μl) was used as the PCR template for 16S rDNA gene amplification. The PCR reaction buffer (total volume, 25 μl) contained 200 μM of each deoxynucleoside triphosphate in 10 mM Tris-HCl (pH 9.0), 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 μM of primers 27F and 1492R[31] and ~2.5 units of puReTaq polymerase as included in the illustra puReTaq Ready-To-Go PCR beads kit (GE Healthcare). The PCR thermal cycling programme was as follows: 95 °C for 2 min; 30 cycles of 95 °C for 30 s, 50 °C for 30 s and 72 °C for 45 s; and 72 °C for 7 min. The PCR products were sent to AGOWA Genomics for unidirectional sequencing using primers 27F and 519R[31]. The quality of the sequences was controlled by removing traces of the sequencing primers by using PHRED[32] with a base-calling score of $n \geq 20$. Ambiguous base calls (that is, 'N's) at the ends of the sequences were also trimmed away. All sequences were analysed using the programs MALLARD[33] and CHECK_CHIMERA from the Ribosomal Database Project[34]. Neither program detected any chimaeras. Resulting sequences were then compared with the SILVA database[35] using the program BLAST[36]. A phylogenetic tree was built to infer the relationships among the ancestral strains and their closest known relatives. In combination with their morphological characteristics, we considered the 95 selected strains as different taxa (Supplementary Information, section 6).

All 95 bacterial strains were tested for production of prophage (that is, to see whether they contained an inducible viral genome) on treatment with and without (control) the inducing agent mitomycin C at a final concentration of 1 μg ml^{-1} (ref 31). Incubations were carried out in 96-well microplates for 24 h. The growth kinetics of each strain was obtained by inoculating 10 μl of overnight cultures in 200 μl marine broth culture medium (BD Difco Marine Broth 2216; autoclaved for 20 min at 121 °C). Cultures were allowed to grow in the dark at 20 °C for 48 h. Changes in cell density were measured by the amount of light absorbance at 590-nm wavelength (FLUOstar OPTIMA spectrophotometer, BMG) every 15 min. We calculated the difference in the absorbance at stationary phase between control and treated samples, and found a bimodal distribution (no effect and strong effect of mitomycin C) with a threshold corresponding to a 30% growth reduction by mitomycin C. We defined a strain containing cells with an inducible viral genome as a strain with a growth reduction larger than this threshold, and therefore consider strains under this threshold to be phage free. Of those strains that fell below the 30% threshold, we randomly selected 31 strains for the selection experiment.

Selection experiment. BIOLOG EcoPlates contain 31 different carbon substrates (plus one blank) belonging to different chemical families. In addition to the carbon substrates, each well contains a fluorogenic tetrazolium dye (5 cyano-2,3 ditolyl tetrazolium chloride), which is reduced to a violet-fluorescent formazan molecule when the carbon source is oxidized. Colour development was measured spectrophotometrically at 590 nm with a FLUOstar OPTIMA spectrophotometer and used as a proxy of metabolic activity[37]. Each of the 31 strains was used to establish three replicates of specialist and generalist selection lines. For the specialist treatment, each strain was assigned at random to one of the 31 carbon substrates. The EcoPlates used for the specialist treatment were prepared 2 h before each transfer with the addition of 140 μl of M9 minimal salts (0.1 g l^{-1} NH_4Cl, 6 g l^{-1} Na_2HPO_4, 3 g l^{-1} KH_2PO_4, 0.5 g l^{-1} NaCl) with salinity adjusted to 35.6 by the addition of NaCl, to match the salinity of the environment from which they were sampled. For the generalist treatment, we obtained a complex medium by mixing all of the 31 carbon EcoPlate substrates. The EcoPlates were prepared as for the specialists and after 30 min the contents of the EcoPlates (except the blanks) were transferred into a sterile flask, mixed with an orbital shaker and redistributed across a 96-well sterile microplate.

Each colony was initially grown for 24 h at 20 °C in 0.5 ml marine broth medium under constant orbital shaking. This solution (10 μl) was then used to inoculate each EcoPlate well from the two selection treatments. We intentionally did not wash the cells of marine broth medium to transfer a small quantity of this medium and assure the survival of the strains. Preliminary trials showed most of the strains did not initially survive in the EcoPlates without the marine broth. Consequently, for the first seven transfers we added a minimal quantity of marine broth, which we reduced from 7% (v/v) to 0% in steps of 1% at each transfer. The bacteria were incubated for 48 h in the dark at 20 °C in humid chambers. Bacteria were serially transferred to maintain maximal growth rate and to renew the substrate. A transfer consisted of inoculating every well from a new plate with 10 μl of the corresponding previous well. Before each transfer, new EcoPlates were prepared as described above. The mixed medium was prepared at every tenth transfer and stocked at 4 °C. The selection experiment ran for 32 transfers (several hundreds generations). At the thirty-second transfer, the contents of every well was amended with glycerol (50% v/v) and frozen at −80 °C.

We measured light absorbance after conducting the selection experiment. Several lineages or replicates went extinct during the experiment, mostly for the specialist treatment. Therefore, we selected the most productive replicate among the lineages that survived the specialist treatment, and the corresponding lineages of the generalist treatment and the ancestors. We used 20 strains/lineages out of the 31 that were subject to the selection experiment for the BEF experiment.

BEF experiment. We assembled random combinations of species at six levels of species richness for each of the three treatments (ancestors, evolved specialists and generalists). We used an experimental design that allowed separation of the effects of species richness and species composition[9,26]. The experimental design consisted of a set of $20/s$ microcosms, each with s species present. Within this set, the microcosm assemblages were constructed by sampling all of the 20 species without replacement. The construction of a system of microcosms was carried out independently four times. We chose values of s to be every factor of 20 ($s = 1, 2, 4, 5, 10, 20$), so for any given s the number of assemblages considered was $4 \times 20/s$. Each assemblage was replicated three times, so in the experiment as a whole there were $3 \times 3 \times 4 \times (1 + 2 + 4 + 5 + 10 + 20) = 1{,}512$ microcosms.

Bacterial communities were assembled in six sterile 96-well, 1-ml microplates. Bacteria were initially grown for 24 h at 20 °C in 6 ml marine broth medium under constant orbital shaking in humid chambers. The cultures were centrifuged (5 min at 3,500 r.p.m.) and washed by eliminating the marine broth and adding 6 ml M9 minimal salts with buffered salinity. Because the cultures had different productivities in the marine broth, we first measured cell density by flow cytometry[38]. We adjusted the cell density to a concentration of 5×10^{-5} cells ml^{-1} with buffered M9. The bacteria were left in starvation for 2 h before 20 μl (40 μl for the monocultures) was inoculated into the appropriate wells. Once the cultures were distributed in the appropriate wells of the six plates, 10 μl of each community was transferred into three replicated microplates containing 140 μl of the mixed medium (that is, the assemblages were initiated with 5×10^{-3} cells), for a total of 18 microplates. The cultures were incubated at 20 °C in humid chambers for 72 h. Light absorbance at 590 nm was measured at 48 and 72 h.

Assays. Assays were conducted to measure strain/lineage performances on each carbon substrate at the end of the experiment. Before the assays, frozen cultures from the end of the selection period were reconditioned in 6 ml marine broth medium for 24 h at 20 °C in humid chambers under constant orbital shaking. The cultures were centrifuged (5 min at 3,500 r.p.m.) and washed by removing the marine broth and adding M9 minimal salts with buffered salinity to adjust cell density to 3.3×10^{-6} cells ml^{-1} (a pilot study showed that this concentration was optimal to obtain a signal differentiating strains). The cultures were left in starvation for 2 h. The EcoPlates were prepared with 120 μl of the buffered M9 solution. The EcoPlates were incubated with 30 μl of culture. Each strain/lineage was incubated in triplicate at 20 °C in humid chambers. Light absorbance at 590 nm was measured after 48 h.

Statistical analyses. The selection treatment (three levels) and species richness (log-transformed) were entered into an analysis of covariance of the bacterial productivity. The dependent variable was the light absorbance at 590 nm and we analysed the effects of the experimental treatments after 48 and 72 h. The results of the assays were averaged over the three replicates. A strain/lineage was considered to be able to exploit a carbon substrate when the light absorbance was larger than the 95% of the distribution of the blanks. For each assemblage, we calculated the NDI, which is the total number of carbon substrates that a community is able to exploit, calculated on the basis of the community composition and the individual ability of each strain/lineage to exploit the carbon substrates. A second analysis of covariance was conducted (excluding the ancestors) considering the NDI instead of species richness. We also fitted a linear model that assessed the effect of species richness and species identity on ecosystem functioning without requiring knowledge of the contribution of individual species to ecosystem functioning in mixture (see ref. 26 for details of this methodology). The model returns species-specific coefficients that could be interpreted as the contribution of individual species to ecosystem functioning relative to the average species.

31. Lane, D. J. in *Nucleic Acid Techniques in Bacterial Systematics* (eds Stackebrandt, E. & Goodfellow, M.) 115–147 (Wiley, 1991).
32. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8,** 186–194 (1998).

33. Ashelford, K. E. *et al.* New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl. Environ. Microbiol.* **72,** 5734–5741 (2006).
34. Maidak, B. L. *et al.* A new version of the RDP (Ribosomal Database Project). *Nucleic Acids Res.* **27,** 171–173 (1999).
35. Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35,** 7188–7196 (2007).
36. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
37. Bochner, B. R. Sleuthing out bacterial identities. *Nature* **339,** 157–158 (1989).
38. Bouvier, T., del Giorgio, P. A. & Gasol, J. M. A comparative study of the cytometric characteristics of high and low nucleic-acid bacterioplankton cells from different aquatic ecosystems. *Environ. Microbiol.* **9,** 2050–2066 (2007).

doi:10.1038/nature09600

Development of asymmetric inhibition underlying direction selectivity in the retina

Wei Wei[1], Aaron M. Hamby[1], Kaili Zhou[1] & Marla B. Feller[1,2]

Establishing precise synaptic connections is crucial to the development of functional neural circuits. The direction-selective circuit in the retina relies upon highly selective wiring of inhibitory inputs from starburst amacrine cells[1] (SACs) onto four subtypes of ON–OFF direction-selective ganglion cells (DSGCs), each preferring motion in one of four cardinal directions[2]. It has been reported in rabbit that the SACs on the 'null' sides of DSGCs form functional GABA (γ-aminobutyric acid)-mediated synapses, whereas those on the preferred sides do not[3]. However, it is not known how the asymmetric wiring between SACs and DSGCs is established during development. Here we report that in transgenic mice with cell-type-specific labelling, the synaptic connections from SACs to DSGCs were of equal strength during the first postnatal week, regardless of whether the SAC was located on the preferred or null side of the DSGC. However, by the end of the second postnatal week, the strength of the synapses made from SACs on the null side of a DSGC significantly increased whereas those made from SACs located on the preferred side remained constant. Blocking retinal activity by intraocular injections of muscimol or gabazine during this period did not alter the development of direction selectivity. Hence, the asymmetric inhibition between the SACs and DSGCs is achieved by a developmental program that specifically strengthens the GABA-mediated inputs from SACs located on the null side, in a manner not dependent on neural activity.

The ability to detect motion in the visual scene is a fundamental computation in the visual system that is first performed in the retina. Motion direction is encoded by DSGCs, which fire a maximum number of action potentials during movement in their preferred direction, but fire minimally for movement in the opposite, or null, direction[4,5]. In the mammalian retina, the directional preference of an ON–OFF DSGC is caused by asymmetric inhibitory inputs: movement in the null direction causes strong inhibition that effectively shunts light-evoked excitatory inputs. Indeed, blocking $GABA_A$ receptors abolishes the directionality of DSGCs by increasing spiking in response to null-direction motion[6–8]. Null-side inhibition is thought to arise from SACs because their processes cofasciculate with DSGC dendrites[9,10], where they form direct GABAergic synapses[3], and because ablation of SACs eliminates the directional preference of DSGCs[11,12].

How SAC–DSGC synapses are organized to provide asymmetric inhibition has been an intriguing but difficult question because no apparent asymmetry is detected in the morphology or the distribution of synaptic markers in DSGCs and SACs[13–15]. The first and only piece of evidence for the synaptic basis of asymmetric inhibition came from a functional study between SAC and DSGC pairs in rabbit retina[3], which suggested that SACs on the null side provide inhibitory inputs to the DSGCs but that those on the preferred side do not. Whether this asymmetric inhibition exists in the mouse is not known. In addition, because the directional preference of an ON–OFF DSGC is present by eye opening[16–18] and the identification of DSGCs and their preferred directions is almost impossible before the onset of the light response, little is known about the developmental program that shapes the SAC–DSGC synapses.

Here we use paired recordings and morphological reconstructions from a double-transgenic mouse line that selectively expresses two variants of green fluorescent protein (GFP) in SACs and nasal-preferring ON–OFF DSGCs (nDSGCs) to characterize the organization and the development of the precise wiring between SACs and DSGCs. These mice were generated by crossing two existing lines: *Drd4–GFP* mice, where *Drd4* promoter-driven GFP expression is restricted to nDSGCs[19], and *mGluR2–GFP* mice (*mGluR2* also known as *Grm2*), where a membrane-tethered human interleukin-2α/GFP fusion protein is expressed specifically in SACs in the retina (Fig. 1a)[20].

To detect functional GABAergic synapses between SACs and DSGCs, we performed targeted whole-cell voltage-clamp recordings from SAC–DSGC pairs in whole-mount retinas. To isolate GABAergic synapses, paired recordings were carried out in the presence of drugs that block excitatory synaptic transmission (Fig. 1b). Alexa dyes were included in the recording pipettes to visualize the dendritic morphology of the recorded pairs (Fig. 1c). Only pairs with overlapping dendritic fields were used for analysis.

Paired recordings were carried out in postnatal-day-4 (P4), P7, P14 and adult mice. At P4, GABAergic currents elicited by SAC depolarization were detected in nDSGCs in 64% of pairs (16 of 25 pairs; Fig. 1b, d), indicating that synapse formation between SACs and nDSGCs occurred before and during the first postnatal week, confirming previous findings[10]. By P7, nearly all pairs showed unitary GABAergic connections (P7: 85%, 29 of 34 pairs), and this high level of connectivity persisted into adulthood (P14–48: 91%, 41 of 45 pairs; Fig. 1d). The evoked response was completely blocked by the $GABA_A$ receptor antagonist gabazine (5 μM, $n = 4$; data not shown), indicating that the GABAergic transmission between SACs and nDSGCs is mediated by $GABA_A$ receptors. We note that the finding that connections were readily detected between SACs located on the preferred side of DSGCs in adult mice is in contrast to previous findings in rabbit[3].

Though SACs located on both the preferred side and the null side formed GABAergic synapses with DSGCs, a significant asymmetry in the unitary synaptic strength emerged along the null-preferred axis during the second postnatal week. Synaptic strength was quantified as the $GABA_A$-receptor-mediated whole-cell conductance. These measurements were restricted to the null-side and preferred-side pairs that had similar amounts of overlap between SAC processes and DSGC dendrites. Unexpectedly, at P4 and P7 the GABAergic conductances from both groups were similar (Fig. 2). However, a significant increase in unitary conductance was detected in the null-side pairs but not in the preferred side pairs in retinas at P14 and older (Fig. 2). Hence, the establishment of the direction-selective circuits is mediated by an asymmetric increase in the strength of the unitary conductance between SACs and DSGCs in the week before eye opening.

The difference in GABAergic conductance from the null- and preferred-side SACs prompted us to examine two possibilities regarding the mechanisms underlying this strengthening. First we tested whether this functional asymmetry was correlated with the number or quality of contacts between SACs and nDSGCs, indicating a preferential adhesion

[1]Department of Molecular & Cell Biology, University of California, Berkeley, California 94720-3200, USA. [2]Helen Wills Neurosciences Institute, University of California, Berkeley, California 94720-3200, USA.

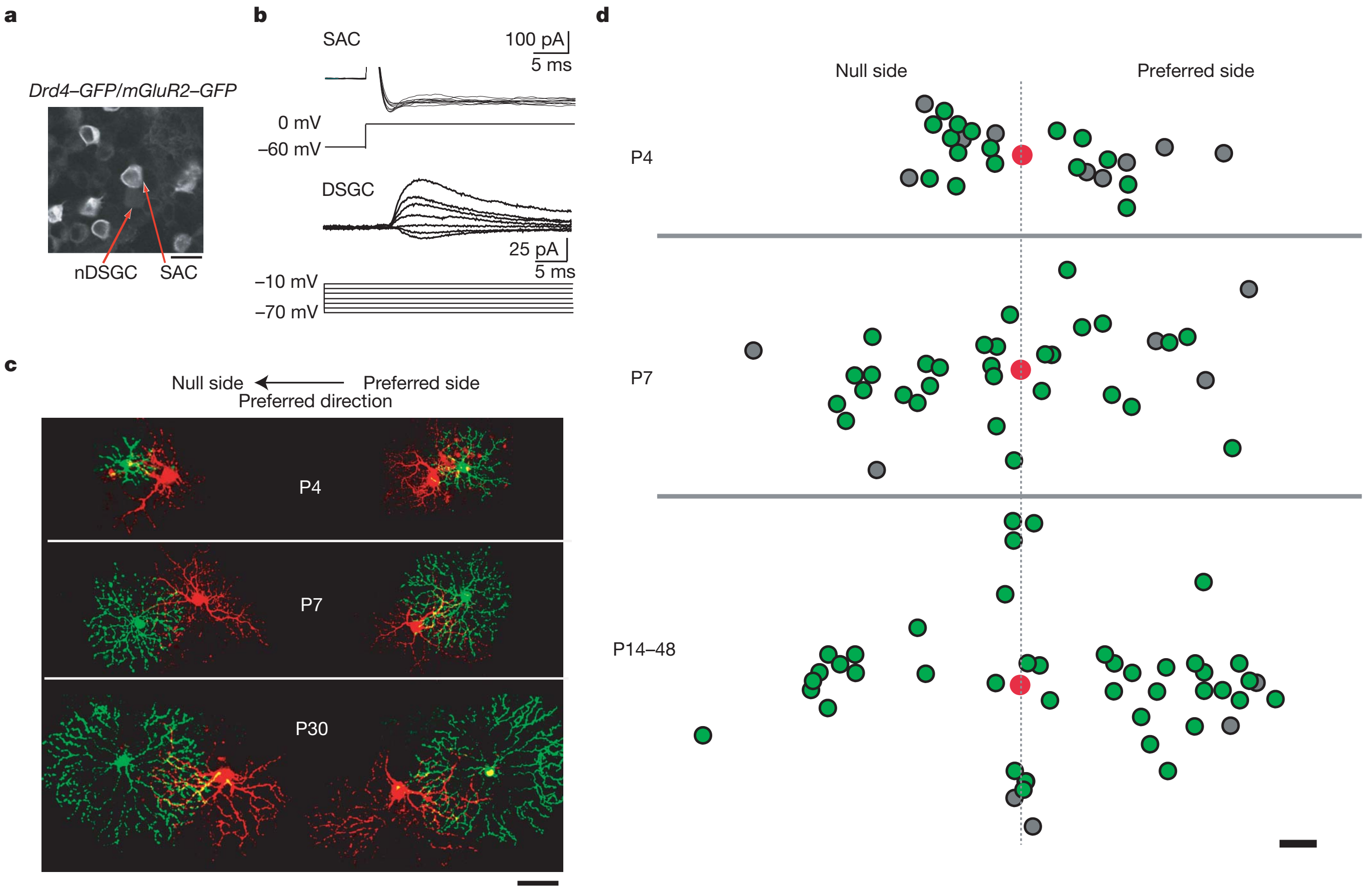


Figure 1 | nDSGCs receive direct GABAergic inputs from SACs located on the null and the preferred side from P4 until adult. a, Fluorescence image of the ganglion cell layer from a P30 *Drd4–GFP/mGluR2–GFP* mouse, showing the bright membrane-bound GFP expressed under the *mGluR2* promoter in the SACs and the dim cytoplasmic GFP driven by the *Drd4* promoter in the nDSGC. Scale bar, 25 μm. **b**, Paired whole-cell voltage-clamp recordings of GABAergic currents in a P4 nDSGC (lower traces) evoked by depolarization of a SAC from the null side (upper traces) in the presence of the NMDA (*N*-methyl-D-aspartate) receptor antagonist D(−)-2-amino-5-phosphonovaleric acid (AP5), α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA)/kainate receptor antagonist 6,7-dinitroquinoxaline-2,3-dione (DNQX) and α4-containing nicotinic acetylcholine receptor antagonist dihydro-β-erythroidine (DHβE). SACs were depolarized from −60 to 0 mV, which reliably evoked an inward current in SACs. The postsynaptic GABAergic currents were recorded in DSGCs at different holding potentials to determine the current–voltage relationship of the conductance. **c**, Example images of synaptically connected, dye-filled SAC–DSGC pairs at P4, P7 and P30. The left-hand side shows pairs with SACs (green) located on the null side of the DSGCs (red). The right-hand side shows preferred-side pairs. Scale bar, 50 μm. **d**, Soma locations of the GABAergically connected SAC–nDSGC pairs along the null-preferred axis during development. Red spots represent the positions of DSGC cell bodies. The positions of SAC cell bodies that form GABAergic synapses with their respective nDSGCs are shown as green spots; the SAC cell bodies that were not connected to nDSGCs are shown as grey spots. All pairs had overlapping dendritic fields. Scale bar, 25 μm.

between SACs located on the null side and DSGCs[3], which is an important mechanism for dendritic differentiation and synaptogenesis in other systems[21,22]. After electrophysiological recording, the dendritic arborizations of the synaptically connected, Alexa-dye-filled SAC–nDSGC pairs from P14 to P48 were imaged live with a two-photon microscope and reconstructed using NEUROLUCIDA (Fig. 3a). We examined the overlapping region between the nDSGC dendrites and the distal portion (roughly the outer third) of the SAC processes enriched in varicosities, which are the sites of neurotransmitter release[9]. Crossing points between distal SAC processes and nDSGC dendrites were defined as 'contacts' (Fig. 3a, inset). A subset of contacts exhibited cofasciculation[9,10,23], which were defined as 2-μm segments along which the processes from the two cells remained in contact (Fig. 3a, inset). The null- and preferred-side SAC–nDSGC pairs showed a similar density of contacts (Fig. 3b) and cofasciculations (Fig. 3c). No asymmetry was found when all of the SAC processes were included in the above analysis (Supplementary Fig. 1). Therefore, the functional asymmetry in GABAergic synapses does not involve selective adhesion between null-side SAC processes and DSGC dendrites[24].

The second possibility we tested was whether spontaneous retinal activity during the second postnatal week has a role in the establishment of direction selectivity. DSGCs are depolarized by retinal waves, and activity could therefore potentially influence the synapse strengthening[25]. To this end, we first confirmed that the GFP-labelled nDSGCs in the *Drd4–GFP* mice showed a clear preference for nasal motion at eye opening (Fig. 4a) that was sensitive to the $GABA_A$ receptor antagonist gabazine (Supplementary Fig. 2), with a direction selectivity index similar to those recorded in the adult (Fig. 4b). We then injected muscimol, a $GABA_A$ receptor agonist, intravitreally into *Drd4–GFP* mice to block all spontaneous and evoked neural activity in the retina[26]. In the presence of muscimol, evoked synaptic transmission from SACs to nDSGCs and spontaneous activity in both cell types were completely suppressed (Fig. 4c and Supplementary Fig. 3a, b). The effectiveness of muscimol injection at blocking activity *in vivo* was confirmed by examining eye-specific segregation of retinogeniculate projections, which is an independent measure of retinal activity (Supplementary Fig. 3c, d), and the persistence of fluorescently labelled muscimol in the retina at 48 h post-injection (Supplementary Fig. 3e).

We assessed the responses of nDSGCs to stationary flashes and drifting gratings in P14–15 mice that had received repeated muscimol injections in the second postnatal week. Muscimol treatment did not prevent the development of direction-selective responses or significantly reduce

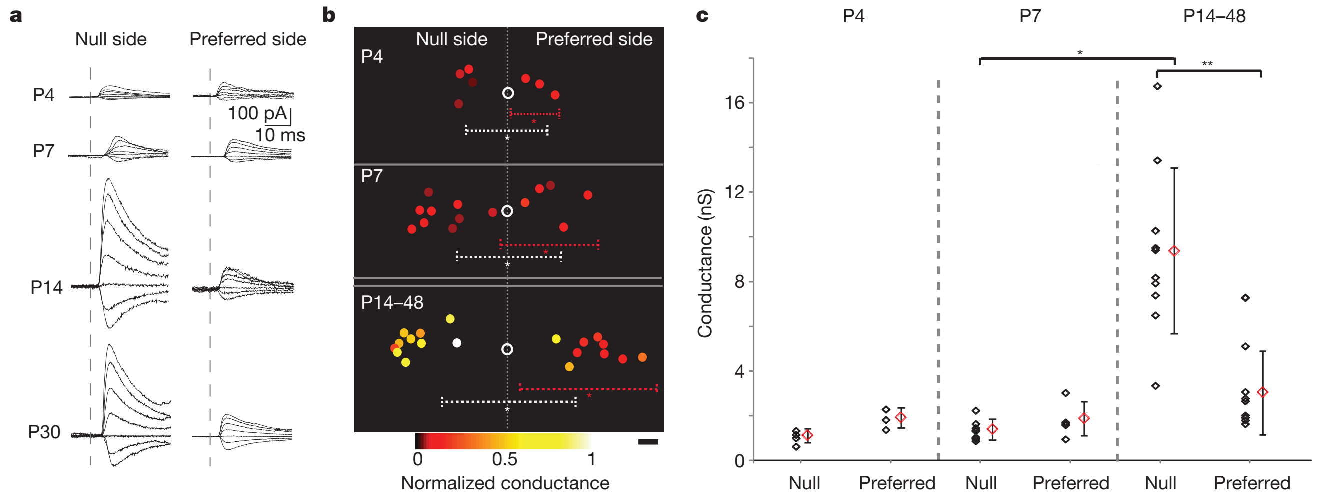


Figure 2 | GABAergic conductance in the null-side SAC–nDSGC pairs strengthens during the second postnatal week. **a**, Postsynaptic GABAergic currents in nDSGCs recorded at holding potentials between −70 and −10 mV in response to depolarization (as in Fig. 1b) of null-side (left) and preferred-side (right) SACs at P4, P7, P14 and P30. **b**, Relative soma positions of SAC–nDSGC pairs used for conductance analysis at P4, P7 and P14–48. Open circles represent nDSGC cell bodies. Filled circles are SAC somas colour-coded for conductance strength normalized to the maximum value across all ages. Dashed lines illustrate average dendritic arborization diameter, centred on the asterisks, for nDSGCs (white) and SACs (red; asterisks represent average soma locations). Scale bar, 25 µm. **c**, Summary plot of GABAergic conductances of the null- and preferred-side SAC–nDSGC pairs at P4, P7 and P14–48. Individual pairs (black) and mean ± s.d. (red) are shown. One-way analysis of variance: $P < 0.0001$; t-test: $*P < 0.0001$, $**P = 0.0003$.

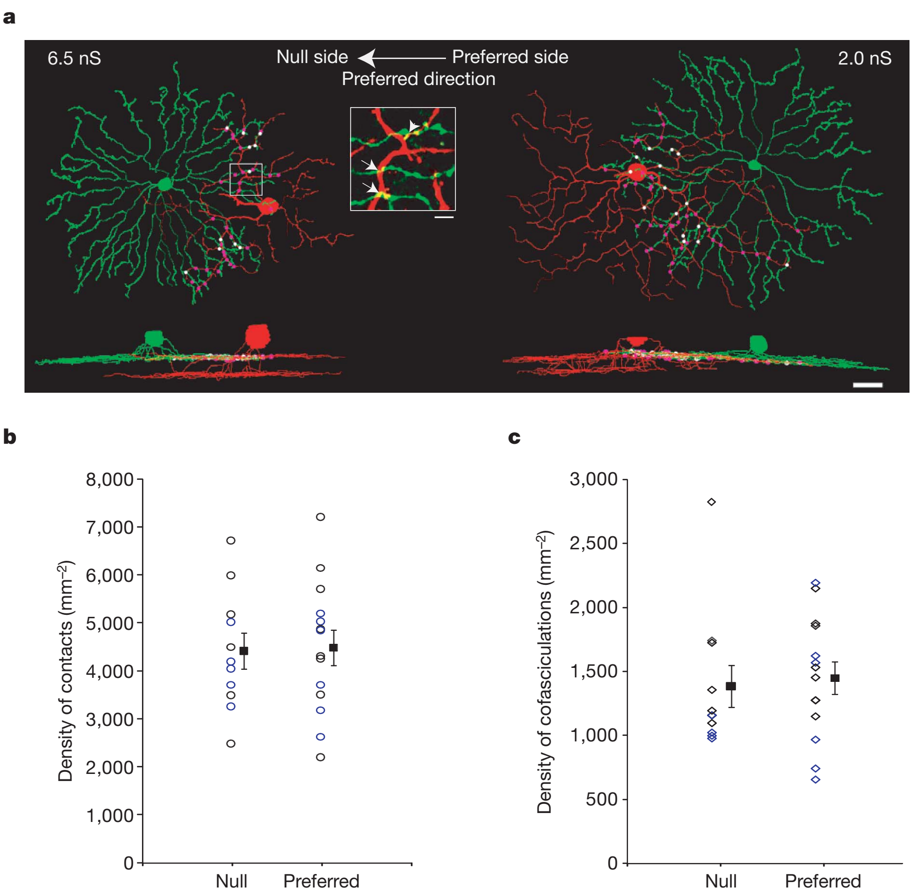


Figure 3 | Dendritic contacts and cofasciculations between SACs and nDSGCs occur at similar densities for the null- and preferred-side pairs. **a**, NEUROLUCIDA reconstructions of the dendrites from the on sublamina and side views of the complete dendritic arborizations from a null-side (left) and a preferred-side (right) pair of SACs and nDSGCs. Dots represent dendritic contacts, with cofasciculation segments coloured white and the rest coloured purple. The GABAergic conductances for the null- and preferred-side pairs are indicated. Scale bar, 25 µm. Inset, fluorescence image of the outlined region showing crossing contacts (arrows) and cofasciculation (arrowhead). Scale bar, 5 µm. **b**, Summary plot of the density of total contacts between DSGCs and distal SAC processes (roughly the outer third) from the null or preferred side from P14 to P48. Individual pairs and mean ± s.d. are shown. The data points for P28 and later are coloured blue, and the ones for before P28 are coloured black. **c**, Summary plot of the density of cofasciculations between nDSGCs and distal SAC processes from the same pairs as in **b**. Null- and preferred-side groups are not significantly different in **b** and **c**. $P > 0.7$, t-test.

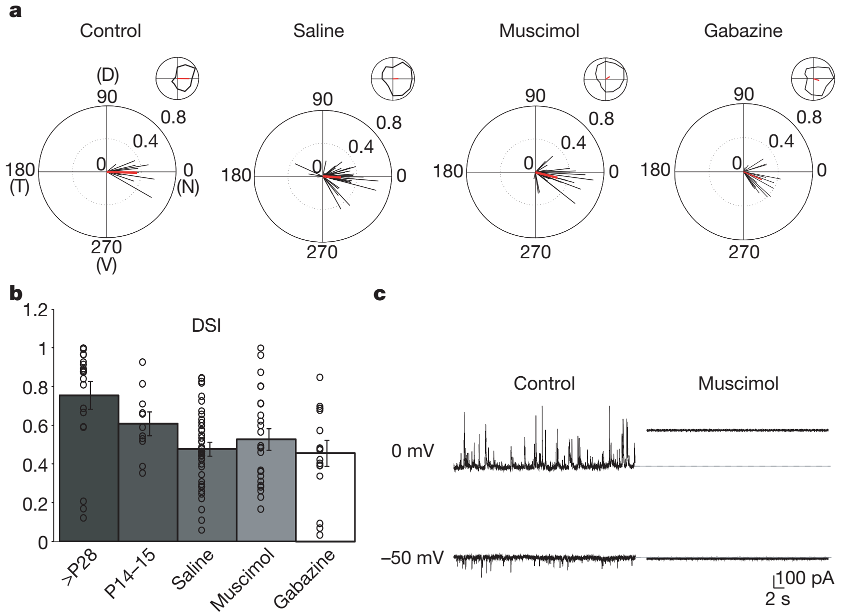


Figure 4 | Intraocular injections of muscimol or gabazine do not alter direction selectivity in nDSGCs. **a**, The normalized spike vector sums of nDSGCs in response to drifting gratings of 12 directions from P14–15 *Drd4–GFP* mice that received either no treatment (control) or intraocular injections of saline, muscimol or gabazine from P6 to P12. D, dorsal; N, nasal; T, temporal; V, ventral. The red lines are mean vector sums of all cells in each group. Insets, examples of normalized tuning curves of single cells, with corresponding vector sums represented as red lines of nDSGCs from each group. Control: $n = 4$ mice, 12 cells; saline: $n = 11$ mice, 43 cells; muscimol: $n = 12$ mice, 25 cells; gabazine: $n = 4$ mice, 17 cells. **b**, Summary plot of direction selectivity index (DSI) for adult (>P28), P14–15 untreated, saline, muscimol and gabazine-treated groups. Bars show mean ± s.e.; open circles represent individual cells. Adult data are reproduced from ref. 19. **c**, Example traces from whole-cell voltage-clamp recordings of inhibitory (upper traces, $V_H = 0$ mV) and excitatory (lower traces, $V_H = -50$ mV) currents from a P14 nDSGC in drug-free artificial cerebrospinal fluid (control, left) or artificial cerebrospinal fluid containing 100 μM muscimol (right). Deflections from baseline correspond to spontaneous synaptic currents. At depolarized potentials, application of muscimol activated a tonic current, which was measured as a change in the baseline holding current[26].

directional tuning of nDSGCs (Fig. 4a, b). Normal ON and OFF light responses were also present in the muscimol-treated group, although there was an increase in the number of cells that did not respond to gratings in the muscimol-treated group (Supplementary Fig. 4).

In visual cortex, activation of $GABA_A$ receptors is required for maturation of GABAergic synapses[27]. To test the hypothesis that $GABA_A$ receptor activation is required for the development of direction selectivity, we performed intravitreal injections of the $GABA_A$ receptor antagonist gabazine into *Drd4–GFP* mice during the second postnatal week. Gabazine treatment did not prevent the development of direction-selective responses of GFP-positive cells to drifting gratings (Fig. 4a, b). Therefore, the development of direction selectivity arises independently of the activation of $GABA_A$ receptors.

To begin exploring the synaptic basis of this increase in conductance between null-side SACs and DSGCs, we recorded the spontaneous inhibitory postsynaptic currents (IPSCs) from nDSGCs at P7 and P14. We found a significant increase in the frequency and no significant change in the amplitude of the GABAergic IPSCs, although there was a trend towards larger IPSC amplitudes at P14 (Supplementary Fig. 5). This result is consistent with the hypothesis that the stronger GABAergic unitary conductance for the null-side pairs is primarily due to increased numbers of functional GABAergic synapses. However, we cannot tell whether the spontaneous IPSCs originate from the null- or preferred-side SACs. Further study is required to determine the relative role increases in synapse number versus synapse strength have in the increase in the unitary conductance between null-side SACs and DSGCs.

Our study demonstrates that asymmetric inhibition arises during the second postnatal week through selective strengthening of the GABAergic conductance from SACs on the null sides of DSGCs. Morphological analysis revealed a similar degree of dendritic contact and cofasciculation between SACs on the null or preferred side, indicating that the synapse development is dissociated from physical encounters between SAC processes and DSGC dendrites, as was recently found in barrel cortex[28].

In addition, we found that blocking depolarization-induced activity or $GABA_A$ receptor activation did not affect the establishment of direction selectivity in the retina, in sharp contrast to direction-selective cells in the visual cortex[29]. This finding lends support to previous studies showing that early visual experience[16–18] or cholinergic retinal waves[18] are not involved in establishing retinal direction selectivity. Therefore, the mechanism underlying the development of retinal direction selectivity is an asymmetric increase in the strength of the inhibitory unitary conductance between SACs and DSGCs in the week before eye opening, without the establishment of asymmetrical dendritic contacts and independent of spontaneous neural activity.

METHODS SUMMARY

We performed dual whole-cell voltage-clamp recordings from SAC–nDSGC pairs in oxygenated artificial cerebrospinal fluid at 32–34 °C containing 119.0 mM NaCl, 26.2 mM $NaHCO_3$, 11 mM glucose, 2.5 mM KCl, 1.0 mM K_2HPO_4, 2.5 mM $CaCl_2$, 1.3 mM $MgCl_2$, 0.05 mM AP5, 0.02 mM DNQX and 0.008 mM DHβE. Recording electrodes of 3–5 MΩ were filled with an internal solution containing 110 mM $CsMeSO_4$, 2.8 mM NaCl, 4 mM EGTA, 5 mM TEA-Cl, 4 mM adenosine 5′-triphosphate (magnesium salt), 0.3 mM guanosine 5′-triphosphate (trisodium salt), 20 mM HEPES and 10 mM phosphocreatine (disodium salt), 0.025 mM Alexa 488 (for SACs) and 0.025 mM Alexa 594 (for nDSGCs), pH 7.2. Data were acquired using PCLAMP 10 recording software and a Multiclamp 700A amplifier (Molecular Devices). The GABAergic conductance was calculated from the linear portion of the current–voltage curve for the SAC-evoked currents in nDSGCs. After recording, we imaged the dye-filled SACs and the nDSGCs using a custom-modified two-photon microscope as described previously[30] (FluoView 300, Olympus America) at 745 nm. Images were acquired at z intervals of 0.5 μm using a ×60 objective (Olympus LUMPlanFl/IR ×60/0.90W). SAC and nDSGC processes were reconstructed from image stacks with NEUROLUCIDA. For *in vivo* injections, we anaesthetized animals with 3.5% isoflurane/2% O_2. The eyelid was then opened with fine forceps, and 1 μl of 10 mM muscimol, 500 μM gabazine or saline was injected using a fine glass micropipette. Injections were made with a picospritzer (World

Precision Instruments) generating 20-p.s.i., 3-ms-long positive pressure. We repeated this procedure every 48 h, starting at P6 and ending at P12. Whole-mount retina preparation and two-photon targeted recording for light responses was performed according to previously described techniques[30]. The direction selectivity index was computed as previously described[19].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 31 May; accepted 13 October 2010.
Published online 5 December 2010.

1. Euler, T., Detwiler, P. B. & Denk, W. Directionally selective calcium signals in dendrites of starburst amacrine cells. *Nature* **418,** 845–852 (2002).
2. Demb, J. B. Cellular mechanisms for direction selectivity in the retina. *Neuron* **55,** 179–186 (2007).
3. Fried, S. I., Munch, T. A. & Werblin, F. S. Mechanisms and circuitry underlying directional selectivity in the retina. *Nature* **420,** 411–414 (2002).
4. Barlow, H. B. & Levick, W. R. The mechanism of directionally selective units in rabbit's retina. *J. Physiol. (Lond.)* **178,** 477–504 (1965).
5. Oyster, C. W. The analysis of image motion by the rabbit retina. *J. Physiol. (Lond.)* **199,** 613–635 (1968).
6. Ariel, M. & Daw, N. W. Pharmacological analysis of directionally sensitive rabbit retinal ganglion cells. *J. Physiol. (Lond.)* **324,** 161–185 (1982).
7. Kittila, C. A. & Massey, S. C. Effect of ON pathway blockade on directional selectivity in the rabbit retina. *J. Neurophysiol.* **73,** 703–712 (1995).
8. Weng, S., Sun, W. & He, S. Identification of ON-OFF direction-selective ganglion cells in the mouse retina. *J. Physiol. (Lond.)* **562,** 915–923 (2005).
9. Famiglietti, E. V. Synaptic organization of starburst amacrine cells in rabbit retina: analysis of serial thin sections by electron microscopy and graphic reconstruction. *J. Comp. Neurol.* **309,** 40–70 (1991).
10. Stacy, R. C. & Wong, R. O. Developmental relationship between cholinergic amacrine cell processes and ganglion cell dendrites of the mouse retina. *J. Comp. Neurol.* **456,** 154–166 (2003).
11. Yoshida, K. *et al.* A key role of starburst amacrine cells in originating retinal directional selectivity and optokinetic eye movement. *Neuron* **30,** 771–780 (2001).
12. Amthor, F. R., Keyser, K. T. & Dmitrieva, N. A. Effects of the destruction of starburst-cholinergic amacrine cells by the toxin AF64A on rabbit retinal directional selectivity. *Vis. Neurosci.* **19,** 495–509 (2002).
13. Chen, Y. C. & Chiao, C. C. Symmetric synaptic patterns between starburst amacrine cells and direction selective ganglion cells in the rabbit retina. *J. Comp. Neurol.* **508,** 175–183 (2008).
14. Famiglietti, E. V. A structural basis for omnidirectional connections between starburst amacrine cells and directionally selective ganglion cells in rabbit retina, with associated bipolar cells. *Vis. Neurosci.* **19,** 145–162 (2002).
15. Jeon, C. J. *et al.* Pattern of synaptic excitation and inhibition upon direction-selective retinal ganglion cells. *J. Comp. Neurol.* **449,** 195–205 (2002).
16. Chan, Y. C. & Chiao, C. C. Effect of visual experience on the maturation of ON-OFF direction selective ganglion cells in the rabbit retina. *Vision Res.* **48,** 2466–2475 (2008).
17. Chen, M., Weng, S., Deng, Q., Xu, Z. & He, S. Physiological properties of direction-selective ganglion cells in early postnatal and adult mouse retina. *J. Physiol. (Lond.)* **587,** 819–828 (2009).
18. Elstrott, J. *et al.* Direction selectivity in the retina is established independent of visual experience and cholinergic retinal waves. *Neuron* **58,** 499–506 (2008).
19. Huberman, A. D. *et al.* Genetic identification of an On-Off direction-selective retinal ganglion cell subtype reveals a layer-specific subcortical map of posterior motion. *Neuron* **62,** 327–334 (2009).
20. Watanabe, D. *et al.* Ablation of cerebellar Golgi cells disrupts synaptic integration involving GABA inhibition and NMDA receptor activation in motor coordination. *Cell* **95,** 17–27 (1998).
21. Togashi, H. *et al.* Cadherin regulates dendritic spine morphogenesis. *Neuron* **35,** 77–89 (2002).
22. Zhu, H. & Luo, L. Diverse functions of N-cadherin in dendritic and axonal terminal arborization of olfactory projection neurons. *Neuron* **42,** 63–75 (2004).
23. Dong, W., Sun, W., Zhang, Y., Chen, X. & He, S. Dendritic relationship between starburst amacrine cells and direction-selective ganglion cells in the rabbit retina. *J. Physiol. (Lond.)* **556,** 11–17 (2004).
24. Vaney, D. I., Collin, S. P. & Young, H. M. in *Neurobiology Of The Inner Retina* (eds Weiler, R. & Osborne, N. N.) 157–168 (Springer, 1989).
25. Elstrott, J. & Feller, M. B. Direction-selective ganglion cells show symmetric participation in retinal waves during development. *J. Neurosci.* **30,** 11197–11201 (2010).
26. Wang, C. T. *et al.* GABA(A) receptor-mediated signaling alters the structure of spontaneous activity in the developing retina. *J. Neurosci.* **27,** 9130–9140 (2007).
27. Huang, Z. J. Activity-dependent development of inhibitory synapses and innervation pattern: role of GABA signalling and beyond. *J. Physiol. (Lond.)* **587,** 1881–1888 (2009).
28. Petreanu, L., Mao, T., Sternson, S. M. & Svoboda, K. The subcellular organization of neocortical excitatory connections. *Nature* **457,** 1142–1145 (2009).
29. Li, Y., Van Hooser, S. D., Mazurek, M., White, L. E. & Fitzpatrick, D. Experience with moving visual stimuli drives the early development of cortical direction selectivity. *Nature* **456,** 952–956 (2008).
30. Wei, W., Elstrott, J. & Feller, M. B. Two-photon targeted recording of GFP-expressing neurons for light responses and live-cell imaging in the mouse retina. *Nature Protocols* **5,** 1347–1352 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Nakanishi for *mGluR2–GFP* mice, A. Huberman for *Drd4–GFP* mice, J. Elstrott for help with MATLAB software, X. Han for mouse genotyping, J. Ledue for imaging assistance and A. Blankenship for reading the manuscript. This work was supported by grants RO1EY013528 and ARRA EY019498 from the National Institutes of Health.

Author Contributions W.W. conducted the electrophysiology and imaging experiments, and manuscript preparation; A.M.H. conducted intraocular injections, analysis of retinogeniculate projection patterns and manuscript preparation. K.Z. conducted NEUROLUCIDA reconstructions and analysis. M.B.F. was involved in the experimental design, data analysis of Supplementary Fig 3c–e and manuscript preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.B.F. (mfeller@berkeley.edu).

METHODS

Mice. *Drd4–GFP* mice in the Swiss Webster background were obtained from MMRRC[19] (http://www.mmrrc.org/strains/231/0231.html), and *mGluR2–GFP* mice were a gift from Shigatada Nakanishi, Osaka. Both strains were backcrossed to the C57BL/6 background in our laboratory. The *Drd4–GFP/mGluR2–GFP* double-transgenic mice were obtained by crossing the two single-transgenic lines.

Whole-cell patch-clamp recording. Single or dual whole-cell voltage-clamp recordings from SACs and nDSGCs were performed in oxygenated artificial cerebrospinal fluid at 32–34 °C containing 119.0 mM NaCl, 26.2 mM $NaHCO_3$, 11 mM glucose, 2.5 mM KCl, 1.0 mM K_2HPO_4, 2.5 mM $CaCl_2$, 1.3 mM $MgCl_2$, 0.05 mM AP5, 0.02 mM DNQX and 0.008 mM DHβE. Recording electrodes of 3–5 MΩ were filled with an internal solution containing 110 mM $CsMeSO_4$, 2.8 mM NaCl, 4 mM EGTA, 5 mM TEA-Cl, 4 mM adenosine 5′-triphosphate (magnesium salt), 0.3 mM guanosine 5′-triphosphate (trisodium salt), 20 mM HEPES and 10 mM phosphocreatine (disodium salt), 0.025 mM Alexa 488 (for SACs) and 0.025 mM Alexa 594 (for nDSGCs), pH 7.25. Data were acquired using PCLAMP 10 recording software and a Multiclamp 700A amplifier (Molecular Devices), filtered at 4 kHz and digitized at a sampling rate of 10 kHz. The GABAergic whole-cell conductance was calculated from the linear portion of the current–voltage curve for the SAC-evoked currents in nDSGCs and analysed using MATLAB software.

Two-photon targeted loose-patch recording of *GFP*-positive neurons for light response. *Drd4–GFP* mice were anaesthetized with isofluorane and decapitated in accordance with the UC Berkeley Institutional Animal Care and Use Committee and in conformance with the NIH Guide for the Care and Use of Laboratory Animals, the Public Health Service Policy and the SFN Policy on the Use of Animals in Neuroscience Research. Under infrared illumination, retinas were isolated from the pigment epithelium in oxygenated Ames' medium (Sigma), cut into dorsal and ventral halves, and mounted over a hole of 1–1.5 mm^2 on filter paper (Millipore) with the photoreceptor layer facing down. Retinas were kept in darkness at 25 °C in Ames' medium bubbled with 95% O2/5% CO2 until use (0–7 h). Recording electrodes of 3–5 MΩ were filled with Ames' medium. GFP fluorescence was detected with a custom-built, FluoView-based two-photon microscope and a Ti:sapphire laser (Coherent) tuned to 920 nm, a wavelength that minimally activates mouse photoreceptors and therefore preserves light response. GFP cells were then targeted for loose-patch recordings using PCLAMP 10 recording software and a Multiclamp 700A amplifier.

Visual stimuli were generated as previously described[19]. Briefly, a white, monochromatic organic light-emitting display (OLEDXL, eMagin; 800 × 600 pixel resolution, 85-Hz refresh rate) was controlled by an Intel Core Duo computer with a Windows XP operating system. Drifting square-wave gratings (spatial frequency, 225 μm per cycle; temporal frequency, 4 cycles s^{-1}; 30° s^{-1} in 12 pseudorandomly chosen directions spaced at 30 intervals, with each presentation lasting 3 s and followed by 500 ms of grey screen) were generated from the OLED using MATLAB and the Psychophysics Toolbox, and were projected through the ×60 water-immersion objective (LUMPlanFl/IR, NA 0.9) via the side port of the microscope, centred on the soma of the recorded cell and focused on the photoreceptor layer. Loose-patch recordings were obtained during the stimulus presentation and analysed using MATLAB. A detailed, step-by-step protocol of the two-photon targeted recording of light response can be found in ref. 30.

Two-photon microscopy and morphological reconstruction. After paired recording, the Alexa-488-filled SACs and the Alexa-594-filled nDSGCs in the *Drd4–GFP/mGluR2–GFP* mice were imaged using the two-photon microscope at 745 nm. At this wavelength, GFP is not efficiently excited but both Alexa 488 and Alexa 594 are brightly fluorescent. Therefore, the morphology of the Alexa-488-filled SACs could be distinguished from the very weak GFP fluorescence. Image stacks were acquired at *z* intervals of 0.5 μm and resampled three times for each stack using a ×60 objective (Olympus LUMPlanFl/IR ×60/0.90W), covering the entire dendritic fields of the SACs and nDSGCs. Image stacks from 25 SAC–nDSGC pairs were then imported into NEUROLUCIDA (MBF Biosciences) and reconstructed in three dimensions. The densities of contacts and cofasciculations were measured from the reconstructions.

Intraocular injections. *Drd4–GFP* animals were anaesthetized with 3.5% isoflurane/2% O_2. The eyelid was then opened with fine forceps, and 1 μl of 10 mM muscimol (Tocris), 500 μM gabazine (Tocris) or saline was injected using a fine glass micropipette. Injections were made with a picospritzer (World Precision Instruments) generating 20-p.s.i., 3-ms-long positive pressure. To prevent efflux of the injected solution, removal of the pipette tip from the eye was done slowly and gentle pressure was then applied to the injection site with a sterile cotton swab for ~10 s. This procedure was repeated every 48 h, starting at P6 and ending at P12.

Statistical analysis. Grouped data are presented as mean ± s.d. or s.e.m. as indicated. Data sets were tested for normality, and statistical differences were examined using one-way analysis of variance and *post hoc* comparisons using Student's *t*-test with Bonferroni corrections (MATLAB).

LETTER

doi:10.1038/nature09602

A high C/O ratio and weak thermal inversion in the atmosphere of exoplanet WASP-12b

Nikku Madhusudhan[1]†, Joseph Harrington[2], Kevin B. Stevenson[2], Sarah Nymeyer[2], Christopher J. Campo[2], Peter J. Wheatley[3], Drake Deming[4], Jasmina Blecic[2], Ryan A. Hardy[2], Nate B. Lust[2], David R. Anderson[5], Andrew Collier-Cameron[6], Christopher B. T. Britt[2], William C. Bowman[2], Leslie Hebb[7], Coel Hellier[5], Pierre F. L. Maxted[5], Don Pollacco[8] & Richard G. West[9]

The carbon-to-oxygen ratio (C/O) in a planet provides critical information about its primordial origins and subsequent evolution. A primordial C/O greater than 0.8 causes a carbide-dominated interior, as opposed to the silicate-dominated composition found on Earth[1]; the atmosphere can also differ from those in the Solar System[1,2]. The solar C/O is 0.54 (ref. 3). Here we report an analysis of dayside multi-wavelength photometry[4,5] of the transiting hot-Jupiter WASP-12b (ref. 6) that reveals C/O ≥ 1 in its atmosphere. The atmosphere is abundant in CO. It is depleted in water vapour and enhanced in methane, each by more than two orders of magnitude compared to a solar-abundance chemical-equilibrium model at the expected temperatures. We also find that the extremely irradiated atmosphere ($T > 2{,}500$ K) of WASP-12b lacks a prominent thermal inversion (or stratosphere) and has very efficient day–night energy circulation. The absence of a strong thermal inversion is in stark contrast to theoretical predictions for the most highly irradiated hot-Jupiter atmospheres[7–9].

The transiting hot Jupiter WASP-12b orbits a star slightly hotter than the Sun (6,300 K) in a circular orbit at a distance of only 0.023 astronomical units (AU), making it one of the hottest exoplanets known[6]. Thermal emission from the dayside atmosphere of WASP-12b has been reported using the Spitzer Space Telescope[10], at 3.6 μm, 4.5 μm, 5.8 μm and 8 μm wavelengths[4], and from ground-based observations in the J (1.2 μm), H (1.6 μm) and Ks (2.1 μm) bands[5] (Fig. 1).

The observations provide constraints on the dayside atmospheric composition and thermal structure, based on the dominant opacity source in each bandpass. The J, H and Ks channels[5] have limited molecular absorption features, and hence probe the deep layers of the planetary atmosphere, at pressure $P \approx 1$ bar, where the temperature $T \approx 3{,}000$ K (Fig. 1). The Spitzer observations[4], on the other hand, are excellent probes of molecular composition. CH_4 has strong absorption features in the 3.6-μm and 8-μm channels, CO has strong absorption in the 4.5-μm channel, and H_2O has its strongest feature in the 5.8-μm channel and weaker features in the 3.6-μm, 4.5-μm and 8-μm channels. The low brightness temperatures in the 3.6-μm (2,700 K) and 4.5-μm (2,500 K) channels, therefore, clearly suggest strong absorption due to CH_4 and CO, respectively. The high brightness temperature in the 5.8-μm channel, on the other hand, indicates low absorption due to H_2O. The strong CO absorption in the 4.5-μm channel also indicates temperature decreasing with altitude, because a thermal inversion would cause emission features of CO in the same channel with a significantly higher flux than at 3.6 μm (refs 11 and 12).

The broadband observations allow us to infer the chemical composition and temperature structure of the dayside atmosphere of WASP-12b using a statistical retrieval technique[11]. We combined a one-dimensional atmosphere model with a Markov-chain Monte Carlo sampler[11,13] that computes over 4×10^6 models to explore the parameter space. The phase space included thermal profiles with and without inversions, and equilibrium and non-equilibrium chemistry over a wide range of atomic abundances. Our models include the dominant sources of infrared opacity in the temperature regime of WASP-12b (refs 14, 15 and 16): H_2O, CO, CH_4, CO_2, H_2–H_2 collision-induced absorption, and TiO and VO where the temperatures

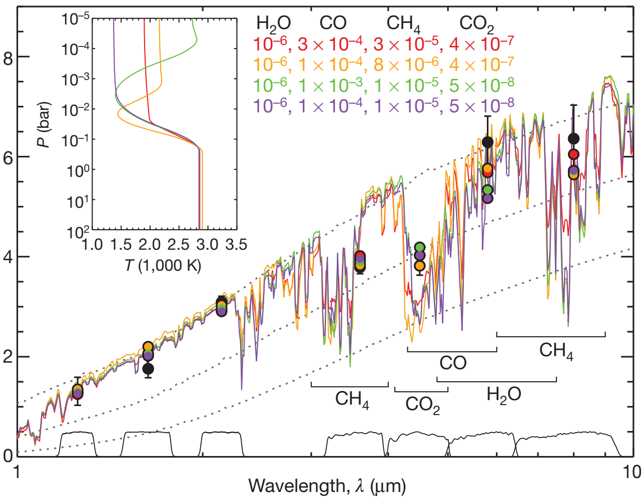


Figure 1 | Observations and model spectra for dayside thermal emission of WASP-12b. F is the flux. The black filled circles with error bars show the data with 1 s.d. errors: four Spitzer observations[4] (3.6 μm, 4.5 μm, 5.8 μm and 8 μm), and three ground-based observations in the J (1.2 μm), H (1.6 μm), and Ks (2.1 μm) bands[5]. Four models fitting the observations are shown in the coloured solid curves in the main panel, and the coloured circles are the channel-integrated model points. The corresponding temperature profiles are shown in the inset. The molecular compositions are shown as number ratio with respect to molecular hydrogen; all the models have C/O between 1 and 1.1. The thin grey dotted lines show the blackbody spectra of WASP-12b at 2,000 K (bottom), 2,500 K (middle) and 3,000 K (top). A Kurucz model[29] was used for the stellar spectrum, assuming uniform illumination over the planetary disk (that is, weighted by 0.5; ref. 7). The black solid lines at the bottom show the photometric band-passes in arbitrary units. The low fluxes at 3.6 μm and 4.5 μm are explained by methane and CO absorption, respectively, required for all the models that fit. The high flux in the 5.8-μm channel indicates less absorption due to H_2O. The observations can be explained to high precision by models without thermal inversions. Models with strong thermal inversions are ruled out by the data (see Fig. 3). The green model features a thermal inversion at low pressures ($P < 0.01$ bar), but the corresponding spectrum is almost indistinguishable from the purple model, which does not have a thermal inversion; both models have identical compositions and identical thermal profiles for $P > 0.01$ bar. Thus, any potential thermal inversion is too weak to be detectable by current instruments.

[1]Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [2]Planetary Sciences Group, Department of Physics, University of Central Florida, Orlando, Florida 32816-2385, USA. [3]Department of Physics, University of Warwick, Coventry, CV4 7AL, UK. [4]NASA's Goddard Space Flight Center, Greenbelt, Maryland 20771-0001, USA. [5]Astrophysics Group, Keele University, Staffordshire ST5 5BG, UK. [6]School of Physics and Astronomy, University of St Andrews, North Haugh, Fife KY16 9SS, UK. [7]Department of Physics and Astronomy, Vanderbilt University, Nashville, Tennessee 37235, USA. [8]Astrophysics Research Centre, School of Mathematics and Physics, Queen's University, University Road, Belfast, BT7 1NN, UK. [9]Department of Physics and Astronomy, University of Leicester, Leicester, LE1 7RH, UK. †Present address: Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey 08544, USA.

are high enough for them to exist in the gas phase[7,17]. The host star has a significantly enhanced metallicity (2 × solar)[6], and evolutionary processes can further enhance the abundances[18,19]; Jupiter has 3 × solar C/H (ref. 18). Our models therefore explore wide abundance ranges: 0.01–100 × solar for C/H and O/H, and 0.1–10 × solar for C/O. Figure 2 shows the mixing ratios of H_2O, CO, CH_4 and CO_2 and the ratios of C/H, O/H and C/O required by the models at different levels of fit. Figure 3 presents the temperature profiles.

We find a surprising lack of water and overabundance of methane (Fig. 2). At 2,000–3,000 K, assuming solar abundances yields CO and H_2O as the dominant species besides H_2 and He (refs 15 and 16). Most of the carbon, and the same amount of oxygen, are present in CO, and some carbon exists as CH_4. The remaining oxygen in a hydrogen-dominated atmosphere is mostly in H_2O; small amounts are also present in species such as CO_2. The CO/H_2 and H_2O/H_2 mixing ratios should each be $>5 \times 10^{-4}$, CH_4/H_2 should be $<10^{-8}$, and CO_2/H_2 should be about 10^{-8}, under equilibrium conditions at a nominal pressure of 0.1 bar. The requirement of $H_2O/H_2 \leq 6 \times 10^{-6}$ and $CH_4/H_2 \geq 8 \times 10^{-6}$ (both at 3σ, 99.73% significance; Fig. 2) is therefore inconsistent with equilibrium chemistry using solar abundances.

The observations place a strict constraint on the C/O ratio. We detect $C/O \geq 1$ at 3σ significance (Fig. 2). Our results rule out a solar C/O of 0.54 at 4.2σ. Our calculations of equilibrium chemistry[16,20] using $C/O = 1$ yield mixing ratios of H_2O, CO and CH_4 that are consistent with the observed constraints. We find that, for $C/O = 1$, H_2O mixing ratios as low as 10^{-7} and CH_4 mixing ratios as high as 10^{-5} can be attained at the 0.1–1 bar level for temperatures around 2,000 K and higher. And, although the CO mixing ratio is predicted to be $>10^{-4}$, making it the dominant molecule after H_2 and He, CO_2 is predicted to be negligible ($<10^{-9}$). The theoretical predictions for a $C/O = 1$ atmosphere are consistent with the observed constraints on H_2O, CH_4, CO and CO_2 (Fig. 2).

The observations rule out a strong thermal inversion deeper than 0.01 bar (Fig. 3). Thermal inversions at lower pressures have opacities too low to induce features in the emission spectrum that current instruments can resolve. For comparison[11,12,21], all stratospheric inversions in Solar System giant planets, and those consistent with hot-Jupiter observations, exist at pressures between 0.01 bar and 1 bar. The major contributions to all the observations come from the lower layers of the atmosphere, $P > 0.01$ bar, where we rule out a thermal inversion (Supplementary Fig. 1). The observations also suggest very efficient day–night energy redistribution (Fig. 2). The low brightness temperatures at 3.6 μm and 4.5 μm imply that only part of the incident stellar energy is re-radiated from the dayside, whereas up to 45% is absorbed and redistributed to the nightside. The possibility of a deep thermal inversion and inefficient redistribution was suggested recently[5], based on observations in the J, H and Ks channels, but the Spitzer observations rule out both conditions.

The lack of a prominent thermal inversion contrasts with recent work that designates WASP-12b as a member of the class of very hot Jupiters that are expected to host inversions[7,22]. At $T > 2{,}000$ K, molecules such as TiO and VO, which are strong absorbers in the ultraviolet/visible, are expected to be in gas phase and potentially cause thermal inversions[7]. WASP-12b, now being the hottest planet without a distinct inversion, presents a major challenge to existing atmospheric classification schemes for exoplanets based on thermal inversions[7,22]. Although there are hints of low chromospheric activity[8] in the host star, it remains to be seen whether the high incident continuum ultraviolet flux expected for WASP-12b might be efficient in photo-dissociating inversion-causing compounds, thus explaining the lack of a strong

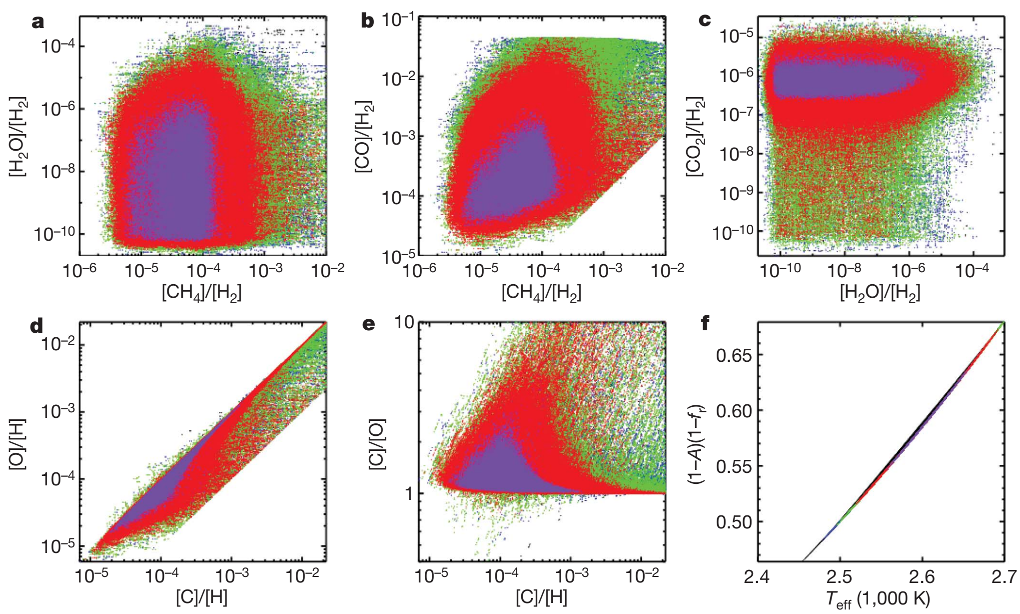


Figure 2 | Constraints on the atmospheric composition of WASP-12b. **a**–**e** The distributions of models fitting the seven observations (Fig. 1) at different levels of χ^2 are shown. The coloured dots show χ^2 surfaces, with each dot representing a model realization. The purple, red, green, blue and black colours correspond to models with χ^2 less than 7, 14, 21 and 28 and $\chi^2 > 28$, respectively (χ^2 ranges between 4.8–51.3). Mixing ratios are shown as ratios by number with respect to H_2. At 3σ significance, the constraints on the composition are $H_2O/H_2 \leq 6 \times 10^{-6}$, $CH_4/H_2 \geq 8 \times 10^{-6}$, $CO/H_2 \geq 6 \times 10^{-5}$, $CO_2/H_2 \leq 5 \times 10^{-6}$, and $C/O > 1$. The compositions of the best-fitting models (with $\chi^2 < 7$) span $H_2O/H_2 = 5 \times 10^{-11}$ to 6×10^{-6}, $CO/H_2 = 3 \times 10^{-5}$ to 3×10^{-3}, $CH_4/H_2 = 4 \times 10^{-6}$ to 8×10^{-4} and $CO_2/H_2 = 2 \times 10^{-7}$ to 7×10^{-6}; the corresponding ranges in C/O and elemental abundances are $C/O = 1$ to 6.6, $C/H = 2 \times 10^{-5}$ to 10^{-3} and $O/H = 2 \times 10^{-5}$ to 10^{-3}. The constraints on the C/H and O/H ratios are governed primarily by the constraints on CO, which is the dominant molecule after H_2 and He. On the basis of thermochemical equilibrium, the inferred CH_4/H_2 and H_2O/H_2 mixing ratios are possible only for $C/O \geq 1$, consistent with our detection of $C/O \geq 1$. **f**, Constraints on the day–night energy redistribution[11], given by $(1 - A)(1 - f_r)$, where A is the bond albedo and f_r is the fraction of incident energy redistributed to the nightside. Up to $f_r = 0.45$ is possible (for $A = 0$). Thus, the observations support very efficient redistribution. An additional observation in the z′ (0.9 μm) band was reported recently[30]. However, the observation implies a value for the orbital eccentricity inconsistent with other data in the literature[4,5]. We therefore decided to exclude this observation from the analysis presented here, although including it does not affect our conclusions regarding the value of C/O or the temperature structure.

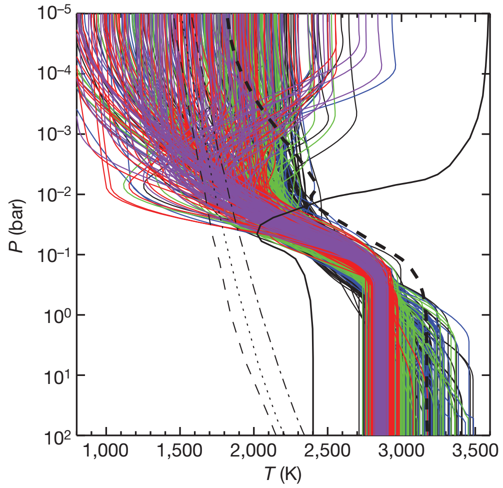


Figure 3 | Thermal profiles of WASP-12b. The solid thin lines show profiles at different degrees of fit (description of colours is as in Fig. 2); only 100 randomly chosen profiles for each χ^2 level are shown, for clarity. The thick black solid curve in the front shows a published profile from a self-consistent model of WASP-12b with a thermal inversion, adapted from ref. 17, which assumes solar abundances. The thick black dashed curve shows the same model but without a thermal inversion. If a thermal inversion is present in WASP-12b, it is expected to be prominent, as shown by the thick solid black curve. A prominent thermal inversion between 0.01 bar and 1 bar is ruled out by the data at 4σ. The ostensibly large inversions in the figure are at low pressures (below 0.01 bar), which have low optical depths, and hence minimal influence on the emergent spectrum (see Fig. 1). The observations are completely consistent with thermal profiles having no inversions. Small thermal inversions are also admissible by the data, and could potentially result from dynamics. The thin black lines show the condensation curves of TiO at solar (dotted), 0.1 × solar (dashed) and 10 × solar (dash-dotted) compositions[17].

inversion[8]. Alternatively, the amount of vertical mixing might be insufficient to keep TiO/VO aloft in the atmosphere to cause thermal inversions[17]. A C/O = 1 might also yield lower TiO/VO than that required to cause a thermal inversion. It is unlikely that the TiO/VO in WASP-12b might be lost to cold traps[17], given the high temperatures in the deep atmosphere on the dayside and nightside.

If high C/O ratios are common, then the formation processes and compositions of extrasolar planets are probably very different from expectations based on Solar System planets. The host star has super-solar metallicity but initial analyses find its C/O consistent with solar[6,23]. In the core accretion model, favoured for the formation of Jupiter, icy planetesimals containing heavy elements coalesce to form the core, followed by gas accretion[19,24]. The abundances of elemental oxygen and carbon are enhanced equally[18,19], maintaining a C/O like the star's. If the host star had a C/O ≈ 1, then the C/O we detect in WASP-12b would have been evident. However, if the stellar C/O is indeed <1, then the C/O enhancement in WASP-12b's atmosphere would suggest either an unexpected origin for the planetesimals, a local overdensity of carbonaceous grains[2,25], or a different formation mechanism entirely. Although carbon-rich giant planets like WASP-12b have not been observed, theory predicts myriad compositions for carbon-dominated solid planets[1,2]. Terrestrial-sized carbon planets, for instance, could be dominated by graphite or diamond interiors, as opposed to the silicate composition of Earth[1,2]. If carbon dominates the heavy elements in the interior of a hot Jupiter, estimates of mass and radius could change compared to those based on solar abundances. Future interior models[26] should investigate the contribution of high C/O to the large radius of WASP-12b: that is, 1.75 Jupiter radii for 1.4 Jupiter masses (ref. 6).

The observed molecular abundances in the dayside atmosphere of WASP-12b motivate the exploration of a new regime in atmospheric chemistry. It remains to be seen whether photochemistry in WASP-12b can significantly alter the composition in the lower layers of the atmosphere at $P = 0.01$–1 bar; these layers contribute most to the observed spectrum (Supplementary Fig. 1). Explaining the observed composition as a result of photochemistry with solar abundances would be challenging. CH_4 is more readily photodissociated than H_2O (refs 9 and 27), and hence a depletion of CH_4 over that predicted with solar abundances might be expected, as opposed to the observed enhancement of CH_4. Apart from the spectroscopically dominant molecules considered in this work, minor species such as OH, C_2H_2 and FeH (refs 27 and 28), which are not detectable by current observations, could potentially be measured with high-resolution spectroscopy in the future. Detection of these species would allow additional constraints on equilibrium and non-equilibrium chemistry in WASP-12b, although their effect on the C/O would be negligible. Models of exoplanetary atmospheres have typically assumed solar abundances and/or solar C/O, thereby exploring a very limited region of parameter space[7,12,20]. Data sufficient for a meaningful constraint on C/O exist for only a few exoplanets. That this initial C/O statistical analysis has C/O ≥ 1 potentially indicates a wide diversity of planetary compositions.

Received 8 July; accepted 21 October 2010.
Published online 8 December 2010.

1. Bond, J. C., O'Brien, D. P. & Lauretta, D. S. The compositional diversity of extrasolar planets. I. *In situ* simulations. *Astrophys. J.* **715,** 1050–1070 (2010).
2. Kuchner, M. & Seager, S. Extrasolar carbon planets. Preprint at ⟨http://arxiv.org/abs/astro-ph/0504214⟩ (2005).
3. Asplund, M., Grevesse, N. & Sauval, A. in. *Cosmic Abundances as Records of Stellar Evolution and Nucleosynthesis* (eds Barnes, T. G. III & Bash, F. N.) 25–38 (ASP Conf. Ser. 336, 2005).
4. Campo, C. *et al.* On the orbit of exoplanet WASP-12b. *Astrophys. J.* (in the press). Preprint at ⟨http://arxiv.org/abs/1003.2763⟩ (2010).
5. Croll, B. *et al.* Near-infrared thermal emission from WASP-12b: detections of the secondary eclipse in Ks, H & J. *Astrophys. J.* (in the press). Preprint at ⟨http://arxiv.org/abs/1009.0071⟩ (2010).
6. Hebb, L. *et al.* 2009, WASP-12b: the hottest transiting extrasolar planet yet discovered. *Astrophys. J.* **693,** 1920–1928 (2009).
7. Fortney, J. J., Lodders, K., Marley, M. S. & Freedman, R. S. A unified theory for the atmospheres of the hot and very hot Jupiters: two classes of irradiated atmospheres. *Astrophys. J.* **678,** 1419–1435 (2008).
8. Knutson, H. A., Howard, A. W. & Isaacson, H. A correlation between stellar activity and hot Jupiter emission spectra. *Astrophys. J.* **720,** 1569–1576 (2010).
9. Zahnle, K. *et al.* Atmospheric sulfur photochemistry on hot Jupiters. *Astrophys. J.* **701,** L20–L24 (2009).
10. Werner, M. W. *et al.* The Spitzer Space Telescope mission. *Astrophys. J. Suppl. Ser.* **154,** 1–9 (2004).
11. Madhusudhan, N. & Seager, S. A temperature and abundance retrieval method for exoplanet atmospheres. *Astrophys. J.* **707,** 24–39 (2009).
12. Burrows, A., Budaj, J. & Hubeny, I. Theoretical spectra and light curves of close-in extrasolar giant planets and comparison with data. *Astrophys. J.* **678,** 1436–1457 (2008).
13. Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (eds) *Markov Chain Monte Carlo in Practice* (Chapman & Hall, 1996).
14. Swain, M. R. *et al.* Molecular signatures in the near-infrared dayside spectrum of HD 189733b. *Astrophys. J.* **690,** L114–L117 (2009).
15. Lodders, K. & Fegley, B. Atmospheric chemistry in giant planets, brown dwarfs, and low-mass dwarf stars. I. Carbon, nitrogen, and oxygen. *Icarus* **155,** 393–424 (2002).
16. Burrows, A. & Sharp, C. M. Chemical equilibrium abundances in brown dwarf and extrasolar giant planet atmospheres. *Astrophys. J.* **512,** 843–863 (1999).
17. Spiegel, D. S., Silverio, K. & Burrows, A. Can TiO explain thermal inversions in the upper atmospheres of irradiated giant planets? *Astrophys. J.* **699,** 1487–1500 (2009).
18. Atreya, S. K. & Wong, A. S. Coupled clouds and chemistry of the giant planets—a case for multiprobes. *Space Sci. Rev.* **116,** 121–136 (2005).
19. Owen, T. *et al.* A low-temperature origin for the planetesimals that formed Jupiter. *Nature* **402,** 269–270 (1999).
20. Seager, S. *et al.* On the dayside thermal emission of hot Jupiters. *Astrophys. J.* **632,** 1122–1131 (2005).
21. Yung, Y. & DeMore, W. B. *Photochemistry of Planetary Atmospheres* (Oxford University Press, 1999).
22. Hubeny, I., Burrows, A., &. Sudarsky, D. A possible bifurcation in atmospheres of strongly irradiated stars and planets. *Astrophys. J.* **594,** 1011–1018 (2003).
23. Fossati, L. *et al.* A detailed spectropolarimetric analysis of the planet-hosting star WASP-12. *Astrophys. J.* **720,** 872–886 (2010).
24. Pollack, J. B. *et al.* Formation of the giant planets by concurrent accretion of solids and gas. *Icarus* **124,** 62–85 (1996).
25. Lodders, K. Jupiter formed with more tar than ice. *Astrophys. J.* **611,** 587–597 (2004).

26. Fortney, J. J., Marley, M. S., &. Barnes, J. W. Planetary radii across five orders of magnitude in mass and stellar insolation: application to transits. *Astrophys. J.* **659,** 1661–1672 (2007).
27. Line, M. R., Liang, M. C. & Yung, Y. L. High-temperature photochemistry in the atmosphere of HD 189733b. *Astrophys. J.* **717,** 496–502 (2010).
28. Cushing, M. C., Rayner, J. T. & Vacca, W. D. An infrared spectroscopic sequence of M, L, and T dwarfs. *Astrophys. J.* **623,** 1115–1140 (2005).
29. Castelli, F. & Kurucz, R. L. New grids of ATLAS9 model atmospheres. Preprint at ⟨http://arxiv.org/abs/astro-ph/0405087⟩ (2004).
30. Lopez-Morales, M. *et al.* Day-side z′-band emission and eccentricity of WASP-12b. *Astrophys. J.* **716,** L36–L40 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the authors of ref. 5 for sharing their ground-based observations before publication, and Thomas J. Loredo for discussions. N.M. thanks S. Seager for financial support during his stay at MIT, where most of the modelling work was carried out. This work is based on observations made with the Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. Support for this work was provided by NASA through an award issued by JPL/Caltech.

Author Contributions N.M. conducted the atmospheric modelling and wrote the paper with input on both from J.H.; J.H. and P.J.W. led the observing proposals, data from which have been interpreted in this work; J.H., J.B. and C.J.C. designed the observations with input from P.J.W., D.R.A., A.C.-C., L.H., C.H., P.F.L.M., D.P. and R.G.W.; J.H., K.B.S., S.N., C.J.C., D.D., J.B., R.A.H., N.B.L., D.R.A., A.C.-C., C.B.T.B. and W.C.B. analysed the Spitzer data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.M. (nmadhu@mit.edu).

doi:10.1038/nature09679

Distributed biological computation with multicellular engineered networks

Sergi Regot[1]*, Javier Macia[2]*, Núria Conde[1,2], Kentaro Furukawa[3], Jimmy Kjellén[3], Tom Peeters[1], Stefan Hohmann[3], Eulàlia de Nadal[1], Francesc Posas[1] & Ricard Solé[2,4,5]

Ongoing efforts within synthetic and systems biology have been directed towards the building of artificial computational devices[1] using engineered biological units as basic building blocks[2,3]. Such efforts, inspired in the standard design of electronic circuits[4–7], are limited by the difficulties arising from wiring the basic computational units (logic gates) through the appropriate connections, each one to be implemented by a different molecule. Here, we show that there is a logically different form of implementing complex Boolean logic computations that reduces wiring constraints thanks to a redundant distribution of the desired output among engineered cells. A practical implementation is presented using a library of engineered yeast cells, which can be combined in multiple ways. Each construct defines a logic function and combining cells and their connections allow building more complex synthetic devices. As a proof of principle, we have implemented many logic functions by using just a few engineered cells. Of note, small modifications and combination of those cells allowed for implementing more complex circuits such as a multiplexer or a 1-bit adder with carry, showing the great potential for re-utilization of small parts of the circuit. Our results support the approach of using cellular consortia as an efficient way of engineering complex tasks not easily solvable using single-cell implementations.

Engineered living cells have been designed to perform a broad variety of functions[8–16] but, with few exceptions, complex computational constructs (such as comparators, bit adders or multiplexers) are difficult to obtain and reuse[17]. Moreover, cell-cell communication requirements rapidly grow with circuit complexity, thus limiting the combinatorial potential of the constructs. One way of overcoming these difficulties is to use cellular consortia[18] based on the idea that external communication between cells in populations involving either single[19–21] or multiple[22–24] cell types would perform functions difficult to be implemented using individual strains. Here, we apply this view to a novel distributed approach based on a reusable, sparse design of synthetic circuits.

A small library of engineered cell types with restricted connections among them was generated, each cell responding to one/two inputs (Fig. 1a–c). The basic two-input and one-output engineered functions include the AND and the inverted IMPLIES (N-IMPLIES, Fig. 1d), which allow implementing any Boolean function. Moreover, some cells define one-input, one-output function (Fig. 1e). The output of each cell type is either a diffusible wiring molecule or the desired output. In contrast with previous works using synthetic consortia[18–22], we have not used cell-cell feedbacks. Instead, cells only respond to an external input and to a single diffusible molecule acting as a wire.

The computation is determined by: (1) the number of cells C involved, (2) the specific function implemented by each engineered cell and (3) the location of cells within the network (see Supplementary Information and Supplementary Fig. 1 for details). Crucially, we allow different engineered cells to produce the output signal, which is thus distributed. Moreover, each cell can be modulated by external inputs, which can either trigger the production of a signal or its inhibition.

The combinatorial nature of our approach is highlighted by calculating, for each C, the number of functions that can be implemented[19]. We have analysed all possible functions with two and three inputs versus C with our approach (see Supplementary Information and Supplementary Fig. 2) and found that most can be constructed using $C = 2–5$ different cells. For instance, in response to three inputs, just three cells results in more than 100 functions and exceed 200 using four cell types. The number of extracellular wires using this approach is significantly lower compared to other standard approximations (Supplementary Fig. 3 and Supplementary Information). With three inputs, over 100 different logical functions can be achieved with only two wires and almost all are obtained with just three to four wires (Supplementary Fig. 3).

As a proof of principle that distributed computation can be implemented *in vivo*, we created a library of engineered yeast cells. Each cell

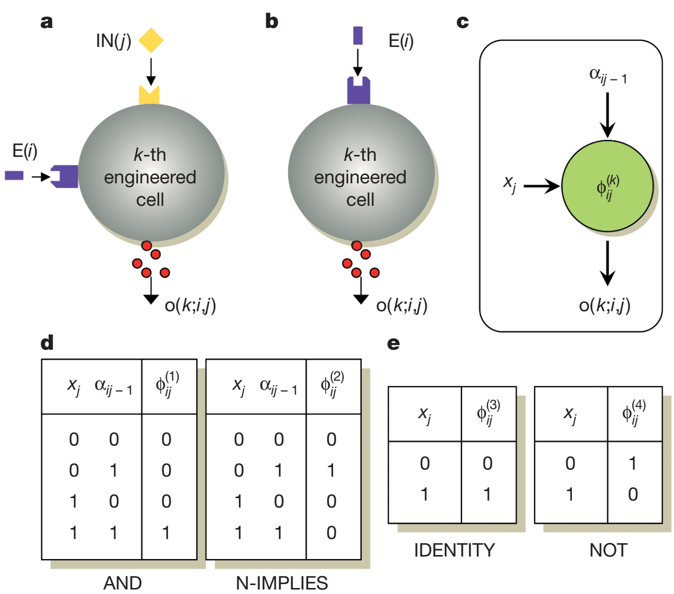


Figure 1 | Basic engineered cells (cell types). a, **b**, Cells can receive signals from other cells (IN) and external sources (E) (**a**) or just from external sources (**b**). Cells can also produce diffusive output molecules. **c**, Representation of the cell behaviour is summarized, where each cell c_{ij} responds to two different inputs; external input (x_j) and a signalling molecule (wire) from another cell (α_{ij-1}). The response of k-th cell type $o(k;i,j)$ can be the production of a new wiring molecule (α_{ij}) or the final output. The k-th cells respond to the presence of signals through some Boolean function $\Phi^{(k)}ij$ with $k = \{1,2,3,4\}$ defining the resulting Boolean output. **d**, **e**, The four basic functions implemented in our study are displayed.

[1]Cell signaling unit, Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra (UPF), E-08003 Barcelona, Spain. [2]ICREA-Complex Systems Laboratory, Universitat Pompeu Fabra (UPF), E-08003 Barcelona, Spain. [3]Department of Cell and Molecular Biology/Microbiology, University of Gothenburg, Box 462, 40530 Gothenburg, Sweden. [4]Santa Fe Institute, Santa Fe, New Mexico 87501, USA. [5]Institut de Biologia Evolutiva, CSIC-UPF, Passeig Maritim de la Barceloneta, 37-49, 08003 Barcelona, Spain.
*These authors contributed equally to this work.

responds to an extracellular stimulus (for example, NaCl, doxycycline, galactose, oestradiol) and/or the presence of a wiring molecule (for example, yeast pheromone). The output of the cells was monitored as the expression of a reporter construct under the control of the *FUS1* promoter (for example, green fluorescent protein, GFP). See Supplementary Fig. 4 for relevant genotype and the logic function of each cell of the library is. The ability of cells to respond to external stimuli (inputs) was monitored by fluorescence in single cell (fluorescence-activated cell sorting, FACS) and normalized to the maximal number of cells able to produce output signal (see Fig. 2a and Supplementary Information). Each cell type has been characterized by its ability to respond to the corresponding stimuli (Fig. 2b and Supplementary Fig. 5).

We then implemented all standard 2-input logic functions by combining just a few engineered cell types. We initially designed a basic circuit with an AND logic (Fig. 3a) involving two cell types responding to two stimuli (NaCl and oestradiol) and using a pheromone (alpha factor) as a wiring molecule. The presence of NaCl stimulates Cell 1 to produce pheromone (IDENTITY) that is received by Cell 2. In addition, Cell 2 has the ability to sense another external input (oestradiol) and it is competent, via the production and activation of the Fus3 mitogen-activated protein kinase (MAPK), to produce the final output. Only in the presence of the two inputs the final outcome was produced (Fig. 3a). Similarly, a NOR gate was implemented using a different pair of cell types in which each cell responded to a particular stimulus (doxycycline and 6a, an inhibitor of Fus3as kinase) with yeast pheromone as a wiring. Only in the absence of both stimuli there was positive output (Fig. 3b).

Next, we designed two completely different circuits that involved the use of three independent engineered cell types, by reusing cells from our previous AND and NOR circuits. The first three-cell circuit was an OR logic gate in which the two inputs are NaCl and galactose. In this circuit, engineered Cell 1 and 5 are IDENTITY functions. They respond to the presence of NaCl (input 1) or galactose (input 2) to produce the wiring molecule that induces output production in Cell 6 (GFP). The presence of any input (galactose or NaCl) generated a positive output as it corresponds to an OR gate (Fig. 3c). Similarly, a NAND gate was designed using doxycycline and glucose as inputs. Cell 3 and Cell 5 display NOT logic. Both secreted pheromone in the absence of stimuli. Cell 6 responded to the presence of pheromone from either Cell 3 or Cell 5 inducing a fluorescent output. As expected, only the presence of both stimuli generates the output (Fig. 3d). This illustrates how to increase computational complexity at low cost. Other circuits can be easily built through reuse (Supplementary Fig. 6). Of note, the N-IMPLIES circuit can be implemented in a single cell (Supplementary Fig. 6a) or by combining cells with different logics (Supplementary Fig. 6b). Using other consortia, we obtained the AND, NOR, OR, NAND, XNOR and XOR gates. However, they can be

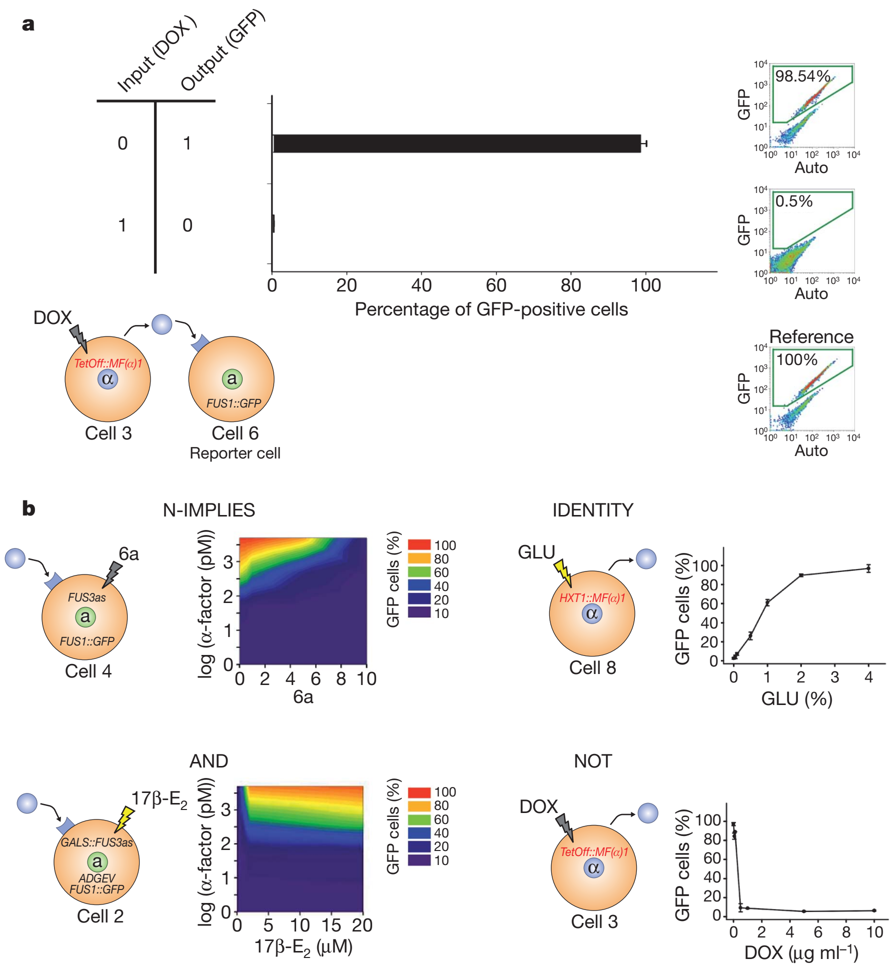


Figure 2 | *In vivo* analyses of engineered cells. a, Quantification of single cell computational output. Truth table and schematic representation of a cell with a NOT logic (see Supplementary Information for complete genotype). The NOT function is implemented in Cell 3, and the reporter cell (Cell 6) is used to quantify alpha factor production *in vivo*. Doxycycline (DOX) was added as indicated and cells were analysed by FACS. Data are expressed as the percentage of GFP-positive cells versus cells treated with pheromone. Results represent the mean ± s.d. of three independent experiments. **b**, Transfer functions of basic logic cells. Schematic representation of cells implementing N-IMPLIES, AND, IDENTITY and NOT functions. Indicated cells were treated with indicated input concentrations (2 inputs, left; 1 input, right). 17β-E$_2$, oestradiol; GLU, glucose.

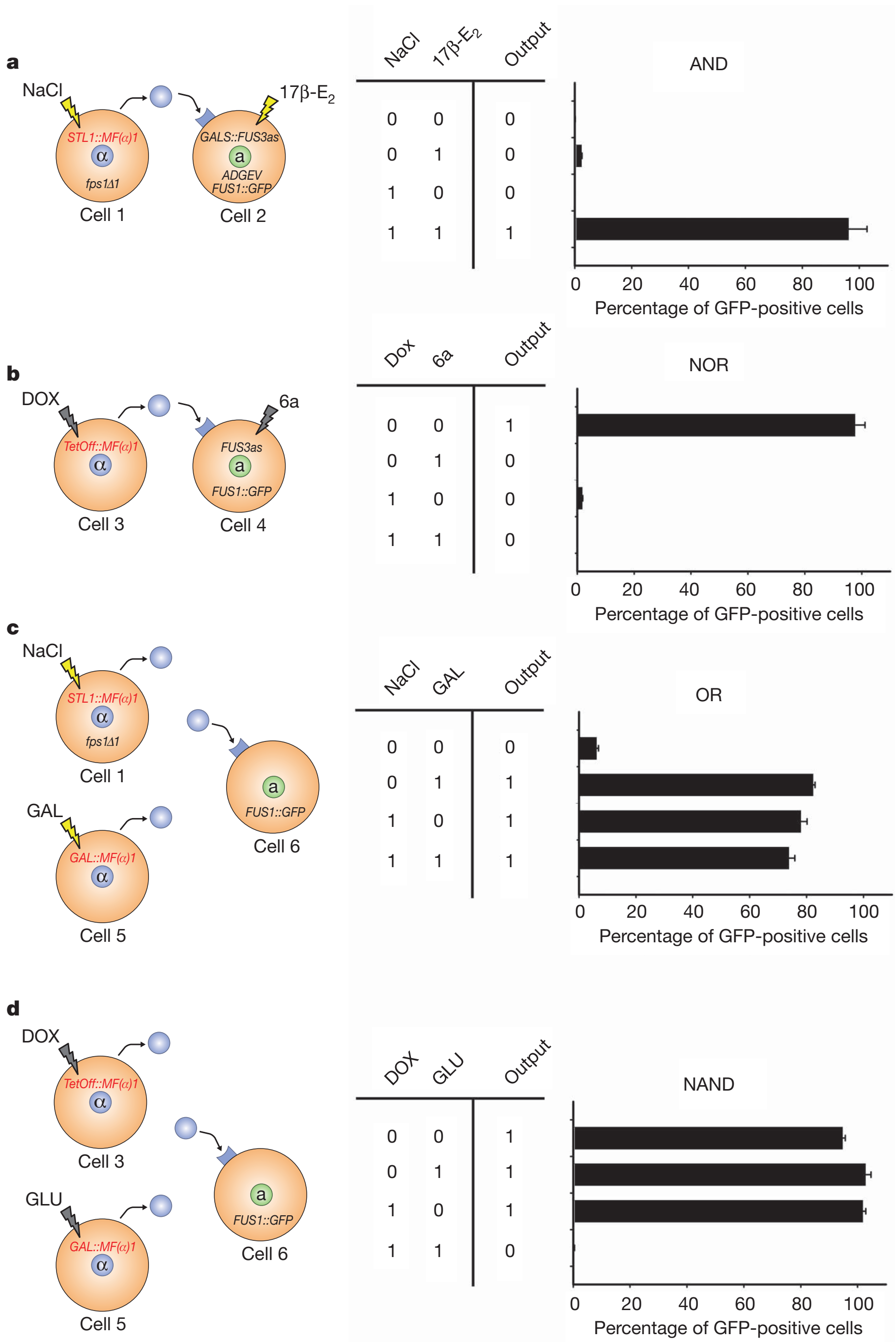


Figure 3 | Engineered cells to implement different logic gates *in vivo*. **a**, Truth table and schematic representation of cells in the AND circuit (see Supplementary Information for complete genotype). Cells were mixed proportionally and inputs (NaCl and oestradiol) were added at the same time. **b**, Panel ordered as in (**a**) following NOR logic. Indicated cells were treated using as inputs doxycycline and 6a. **c**, OR gate. Indicated cells were treated using as inputs 0.4 M NaCl and 2% galactose (GAL). **d**, NAND gate. Indicated strains were treated using as inputs doxycycline and 2% glucose. Data represent the mean and standard deviation of three independent experiments.

implemented using the same inputs for all of them (that is, doxycycline and glucose) (Supplementary Fig. 7a–f). This supports that our approach is adaptable and that multiple functions can be constructed from a small library of reusable cells.

We then analysed the long-term dynamical response of our AND circuit under changing inputs (Supplementary Fig. 7a). We have found that once the circuit is turned on, it can maintain maximal signal for periods beyond 9 h in the presence of stimuli (Supplementary Fig. 8a). Furthermore, once the system has been established, it responds equally well (at least eight generations) while the culture is maintained to a log phase (Supplementary Fig. 8b). We have also experimentally addressed network responsiveness to dynamic changes by means of a microfluidic device containing the cellular types of the AND circuit and have exposed it to changes in the input signals over time. The system was able to dynamically respond to re-stimulation after GFP inactivation (Supplementary Fig. 8c).

Our system can be selectively switched off and partially reprogrammed. For instance, the inhibition of the intracellular signal transduction in Cell 2 of the AND gate, blocks the positive outcome of the circuit (Supplementary Fig. 9a). More interestingly, when reprogramming is applied to complex multicellular circuits, different computations can be obtained with little effort. For instance, when a reprogramming

molecule (glucose) is added to the OR gate shown in Supplementary Fig. 9b it works as IDENTITY for NaCl (input 2). In this context, despite the fact that multicellular circuits might seem more complex, their easy reuse and combination actually makes them more appropriate in many situations.

Finally, we were able to engineer complex circuits by re-using our previous designs. One of them is the multiplexer MUX2to1 circuit that selects one of different input signals and forwards the selected input into a single output. This circuit, if designed in a single cell would be difficult to implement *in vivo* (see Supplementary Fig. 10a). However, using distributed computation, the circuit can be assembled from just three engineered cell types responding to three input signals and a single wiring molecule (Supplementary Fig. 10b). In addition, we also implemented a MUX2to1 circuit that contains four cell types but uses two independent wiring molecules (α-factor from *Saccharomyces cerevisiae* and the α-factor from *Candida albicans).* Cell 10 and Cell 13 respond to doxycycline and produce each one of the wiring molecules. Cell 12 responds to oestradiol and *S. cerevisiae* pheromone whereas Cell 15 responds to galactose and *C. albicans* pheromone. The final output (GFP) is generated by Cell 12 and Cell 15. Here, although the complexity of the circuit required a differential output to eight different input combinations, the *in vivo* results clearly showed that the computation of the three inputs yielded the expected response (Fig. 4a). A second complex circuit, the 1-bit adder with carry, was built by combining XOR and AND gates that respond to the same input (doxycycline and glucose) with two wiring molecules (α-factor from *S. cerevisiae* and *C. albicans*). In addition, output cells express different reporter proteins, a green reporter (adder) or a red reporter (carry) (*FUS1::GFP* or *FUS1::mCherry* respectively) allowing to detect the outcome of the carry and adder in the same culture. The system responds as an XOR gate (green columns) but presence of the two stimuli induces led to a 1-bit carrier (red columns) (Fig. 4b).

Possible applications as well as some caveats will need to be addressed in future work (such as scalability[25,26], strategies for reducing

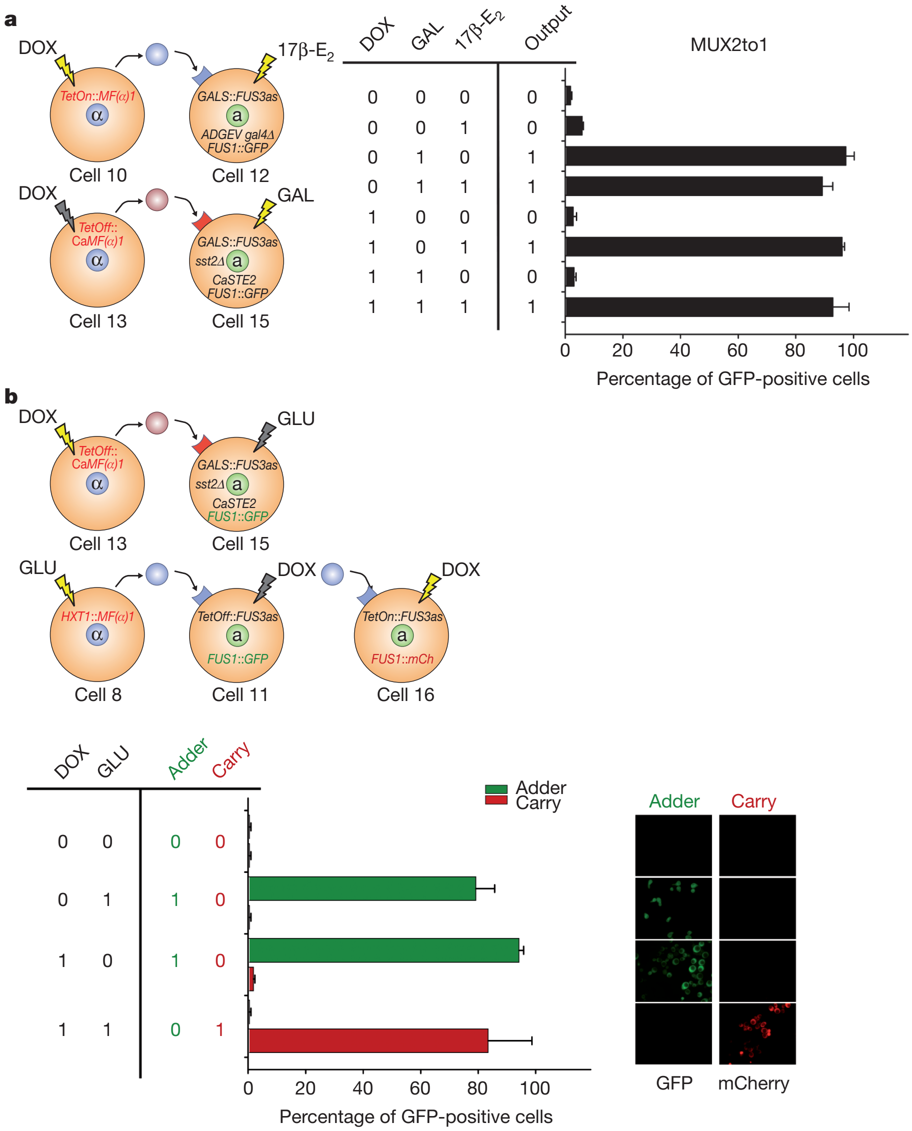


Figure 4 | Design and *in vivo* implementation of a multiplexer (MUX2to1) and 1-bit adder with carry. **a**, Truth table and schematic representation of the cells used in the MUX2to1. Indicated cells were treated using doxycycline (selector) and the inputs oestradiol and/or 2% galactose. Data are expressed as the percentage of GFP-positive cells using a sample treated with either *S. cerevisiae* or *C. albicans* alpha factor as a reference for Cell 12 or Cell 15, respectively. **b**, Truth table and schematic representation of cells used for 1-bit adder with carry. Four cells with two wiring systems that respond to glucose and doxycycline with an XOR logic were combined with an extra cell that respond to same stimuli but with an AND logic in which instead of GFP, mCherry was expressed as output. The final outcome was measured as in Fig. 3a. Green bars indicate the adder output (GFP) whereas red bars represent the carry bit (mCherry). GFP and mCherry images of cells are shown (right panels). Data represent the mean and standard deviation of three independent experiments.

potential crosstalk[3] and robustness to noise[27]). However, the results reported here show that distributed computation using consortia is a powerful strategy to build complex synthetic constructs, opening a new door to reusable, reprogrammable complex circuits (Supplementary Fig. 11).

METHODS SUMMARY

Complete list of engineered yeast strains and plasmids is described in Supplementary Information. Computational output detection was done in single cells by flow cytometry (FACScalibur, Becton Dickinson) and the dynamical inputs responses to different circuits were analysed in a microscopy based microfluidic platform. See Supplementary Information for details about experimental methods.

Received 18 February; accepted 18 November 2010.

Published online 8 December 2010.

1. Bashor, C. J. & Horwitz, A. A. Peisajovich , S. G. , & Lim, W. A. Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living. *Ann. Rev. Biophys.* **39,** 515–537 (2010).
2. Amos, M. *Cellular Computing* (Oxford University Press, 2004).
3. Andrianantoandro, E., Basu, S., Karig, D. & Weiss, R. Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.* **2,** E1–E14 (2006).
4. Bray, D. Protein molecules as computational elements in living cells. *Nature* **376,** 307–312 (1995).
5. Hjelmfelt, A., Weinbeger, E. D. & Ross, J. Chemical implementation of neural networks and Turing machines. *Proc. Natl Acad. Sci. USA* **88,** 10983–10987 (1991).
6. McAdams, H. H. & Shapiro, L. Circuit simulation of genetic circuits. *Science* **269,** 650–656 (1995).
7. Arkin, A. & Ross, J. Computational functions in biochemical reaction networks. *Biophys. J.* **67,** 560–578 (1994).
8. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403,** 335–338 (2000).
9. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403,** 339–342 (2000).
10. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405,** 590–593 (2000).
11. Voigt, C. A. Genetic parts to program bacteria. *Curr. Opin. Biotechnol.* **17,** 548–557 (2006).
12. Kramer, B. P., Fischer, C. & Fussenegger, M. BioLogic gates enable logical transcription control in mammalian cells. *Biotechnol. Bioeng.* **87,** 478–484 (2004).
13. Win, M. N. & Smolke, C. D. Higher-order cellular information processing with synthetic RNA devices. *Science* **322,** 456–460 (2008).
14. Greber, D. & Fussenegger, M. An engineered mammalian band-pass network. *Nucleic Acids Res.* **38,** e174 (2010).
15. Weber, W. *et al.* A synthetic mammalian gene circuit reveals antituberculosis compounds. *Proc. Natl Acad. Sci. USA* **105,** 9994–9998 (2008).
16. Rinaudo, K. *et al.* A universal RNAi-based logic evaluator that operates in mammalian cells. *Nature Biotechnol.* **25,** 795–801 (2007).
17. Purnick, P. E. M. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nature Rev. Mol. Cell Biol.* **10,** 410–422 (2009).
18. Brenner, K., You, L. & Arnold, F. H. Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol.* **26,** 483–489 (2008).
19. Brenner, K., Karig, D. K., Weiss, R. & Arnold, F. H. Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium. *Proc. Natl Acad. Sci. USA* **104,** 17300–17304 (2007).
20. Danino, T., Mondragón-Palomino, O., Tsimring, L. & Hasty, J. A synchronized quorum of genetic clocks. *Nature* **463,** 326–330 (2010).
21. You, L., Cox, R. S. III, Weiss, R. & Arnold, F. H. Programmed population control by cell-cell communication and regulated killing. *Nature* **428,** 868–871 (2004).
22. Shou, W., Ram, S. & Vilar, J. M. G. Synthetic cooperation in engineered yeast populations. *Proc. Natl Acad. Sci. USA* **104,** 1877–1882 (2007).
23. Weber, W., Daoud-El Baba, M. & Fussenegger, M. Synthetic ecosystems based on airbone inter- and intrakingdom communication. *Proc. Natl Acad. Sci. USA* **104,** 10435–10440 (2007).
24. Balagaddé, F. K. *et al.* A synthetic *Escherichia coli* predator–prey ecosystem. *Mol. Syst. Biol.* **4,** 187 (2008).
25. Lucks, J. B., Qi, L., Whitaker, W. R. & Arkin, A. P. Toward scalable parts families for predictable design of biological circuits. *Curr. Opin. Microbiol.* **11,** 567–573 (2008).
26. Chin, J. W. Programming and engineering biological networks. *Curr. Opin. Struct. Biol.* **16,** 551–556 (2006).
27. Rao, C. V., Wolf, D. M. & Arkin, A. P. Control, exploitation and tolerance of intracellular noise. *Nature* **420,** 231–237 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Subirana and S. Ovejas for technical support and S. Pellet and M. Peter for their help in setting up the microfluidics platform, and K. Kuchler for the *FUS1*-mCherry construct. S.R. is recipient of a FPU fellowship (Spanish Government). This work was supported by grants from the James McDonnell Foundation to R.S., the MICINN (BIO2009-07762 and FIS2009-12365); Consolider Ingenio 2010 programme (grant CSD2007-0015), from the ESF (ERAS-CT-2003-980409) FP6 as part of a EURYI scheme award (www.esf.org/euryi) to F.P. and the CELLCOMPUT (FP6) project to F.P., R.S. and S.H., and FP7 UNICELLSYS grant (#201142) to F.P. and S.H., and The Santa Fe Institute to R.S.; F.P. and R.S. laboratories are also supported by the Fundación Marcelino Botín (FMB). F.P. is recipient of an ICREA Acadèmia (Generalitat de Catalunya).

Author Contributions All authors shared all the phases of the work. J.M. and R.S. developed the theoretical background for multicellular computing. Circuits were designed by S.R., J.M., N.C., E.N., F.P. and R.S.; S.R., N.C., K.F., J.K. and T.P. did the experimental designs. J.M., F.P., E.N., S.H. and R.S. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.S. (ricard.sole@upf.edu) or F.P. (francesc.posas@upf.edu).

LETTER

doi:10.1038/nature09684

Images of a fourth planet orbiting HR 8799

Christian Marois[1], B. Zuckerman[2], Quinn M. Konopacky[3], Bruce Macintosh[3] & Travis Barman[4]

High-contrast near-infrared imaging of the nearby star HR 8799 has shown three giant planets[1]. Such images were possible because of the wide orbits (>25 astronomical units, where 1 AU is the Earth–Sun distance) and youth (<100 Myr) of the imaged planets, which are still hot and bright as they radiate away gravitational energy acquired during their formation. An important area of contention in the exoplanet community is whether outer planets (>10 AU) more massive than Jupiter form by way of one-step gravitational instabilities[2] or, rather, through a two-step process involving accretion of a core followed by accumulation of a massive outer envelope composed primarily of hydrogen and helium[3]. Here we report the presence of a fourth planet, interior to and of about the same mass as the other three. The system, with this additional planet, represents a challenge for current planet formation models as none of them can explain the *in situ* formation of all four planets. With its four young giant planets and known cold/warm debris belts[4], the HR 8799 planetary system is a unique laboratory in which to study the formation and evolution of giant planets at wide (>10 AU) separations.

New near-infrared observations of HR 8799, optimized for detecting close-in planets, were made at the Keck II telescope in 2009 and 2010. (See Table 1 for a summary.) A subset of the images is presented in Fig. 1. A fourth planet, designated HR 8799e, is detected at six different epochs at an averaged projected separation of $0.368'' \pm 0.003$” (14.5 ± 0.4 AU). Planet e is bound to the star and is orbiting anticlockwise (see Fig. 2), as are the three other known planets in the system. The measured orbital motion, 46 ± 10 mas yr^{-1}, is consistent with a roughly circular orbit of semimajor axis (a) 14.5 AU with a ~50-year period.

Knowledge of the age and luminosity of the planets is critical for deriving their fundamental properties, including mass. In 2008 we used various techniques to estimate an age of 60 Myr with a plausible range between 30 and 160 Myr (here we represent this as 60^{+100}_{-30} Myr), consistent with an earlier estimate of 20–150 Myr (ref. 5). Two recent analyses (R. Doyon *et al.*, and B. Zuckerman *et al.*, manuscripts in preparation) independently deduce that HR 8799 is very likely to be a member of the 30 Myr Columba association[6]. This conclusion is based on common Galactic space motions and age indicators for stars located between the previously-known Columba members and HR 8799. The younger age suggests smaller planet masses, but to be conservative, we use both age ranges (30^{+20}_{-10} Myr (Columba association) and 60^{+100}_{-30} Myr[1]) to derive the physical properties of planet e.

Table 1 | HR 8799e astrometry, photometry and physical characteristics

Epoch, band, wavelength	Separation [E, N] from the host star
2009 Jul. 31, Kp band 2.124 μm (±0.019″)	[−0.299″, −0.217″]
2009 Aug. 1, L′ band 3.776 μm (±0.013″)	[−0.303″, −0.209″]
2009 Nov. 1, L′ band 3.776 μm (±0.010″)	[−0.304″, −0.196″]
2010 Jul. 13, Ks band 2.146 μm (±0.008″)	[−0.325″, −0.173″]
2010 Jul. 21, L′ band 3.776 μm (±0.011″)	[−0.324″, −0.175″]
2010 Oct. 30, L′ band 3.776 μm (±0.010″)	[−0.334″, −0.162″]
Parameter	**Value**
Projected separation, avg. from all epochs* (AU)	14.5 ± 0.4
Orbital motion (arcsec yr^{-1})	0.046 ± 0.010
Period for a face-on circular orbit (yr)	~50
ΔKs 2.146 μm† (mag)	10.67 ± 0.22
ΔL′ 3.776 μm† (mag)	9.37 ± 0.12
Absolute magnitude at 2.146 μm, M_{Ks} (mag)	12.93 ± 0.22
Absolute magnitude at 3.776 μm, $M_{L'}$ (mag)	11.61 ± 0.12
Luminosity (log $L_\odot$)	-4.7 ± 0.2
Mass for 30^{+20}_{-10} Myr (M_{Jup})	7^{+3}_{-2}
Mass for 60^{+100}_{-30} Myr (M_{Jup})	10^{+3}_{-3}

* The projected separation error (in AU) also accounts for the uncertainty in the distance to the star.
† Planet-to-star flux ratios, expressed as difference of magnitude. No reliable photometry was derived for the Kp-band 2009 Jul. 31 data.

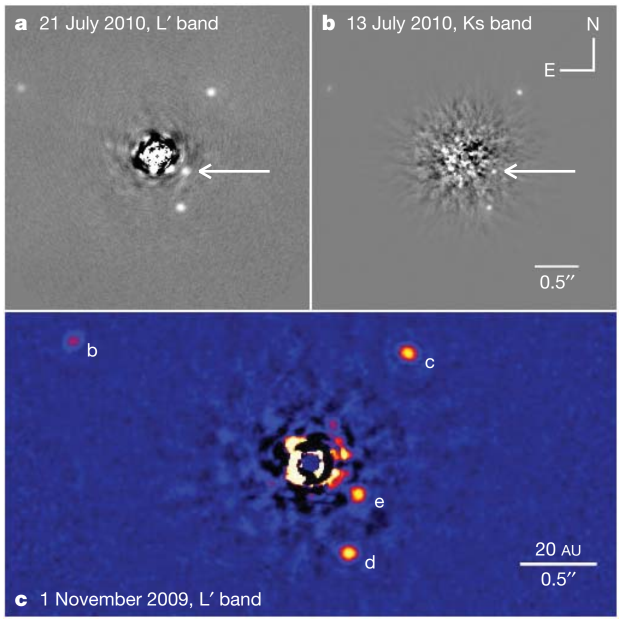


Figure 1 | HR 8799e discovery images. Images of HR 8799 (a star at 39.4 ± 1.0 pc and located in the Pegasus constellation) were acquired at the Keck II telescope with the Angular Differential Imaging technique (ADI)[22] to allow a stable quasi-static point spread function (PSF) while leaving the field-of-view to rotate with time while tracking the star in the sky. The ADI/LOCI[22,23] SOSIE software[24] was used to subtract the stellar flux, and to combine and flux-calibrate the images. Our SOSIE software[24] iteratively fits the planet PSF to derive relative astrometry and photometry (the star position and its photometry were obtained from unsaturated data or from its PSF core that was detectable through a flux-calibrated focal plane mask). **a**, An L′-band image acquired on 21 July 2010; **b**, a Ks-band image acquired on 13 July 2010 (arrows in **a** and **b** point towards planet e); **c**, an L′-band image acquired on 1 November 2009. All three sequences were ~1 h long. No coronagraphic focal plane mask was used on 1 November 2009, but a 400-mas-diameter mask was used on 13 July and 21 July 2010. HR 8799e is located southwest of the star. Planets b, c and d are seen at respective projected separations of 68, 38 and 24 AU from the central star, consistent with roughly circular orbits at inclinations of <40° (refs 11–13). Their masses (7, 10 and 10 M_{Jup} for b, c and d for 60 Myr age[1]; 5, 7 and 7 M_{Jup} for 30 Myr age) were estimated from their luminosities using age-dependent evolutionary models[25]. North is up and east is left.

[1]National Research Council Canada, Herzberg Institute of Astrophysics, 5071 West Saanich Road, Victoria, British Columbia V9E 2E7, Canada. [2]Physics & Astronomy Department, University of California, Los Angeles, California 90095, USA. [3]Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, California 94550, USA. [4]Lowell Observatory, 1400 West Mars Hill Road, Flagstaff, Arizona 86001, USA.

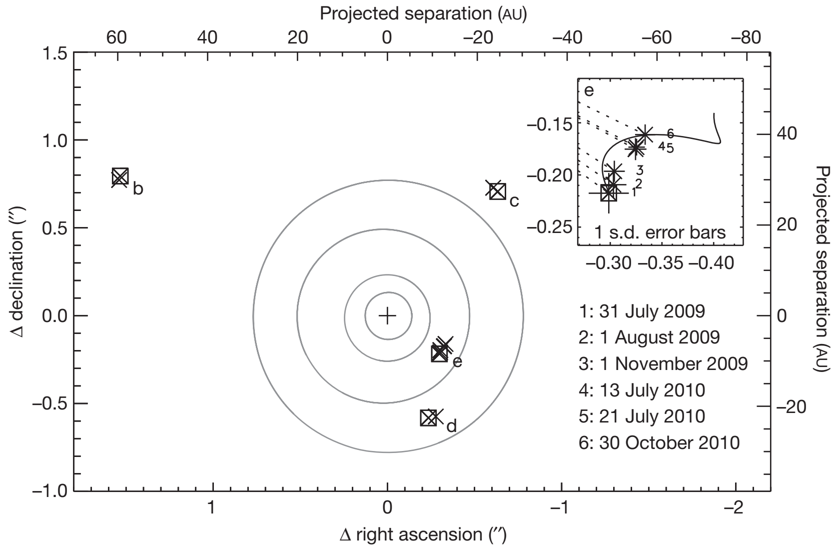


Figure 2 | HR 8799e 2009–10 astrometry. Main figure, the 2009–10 orbital motions of the four planets—b, c, d, e. Crosses denote the positions for 2009 and 2010 first an last epochs for b, c and d, and for all six epochs for e. A square is drawn over the cross symbol of each planet's first epoch. Inset, a zoomed version of planet e's astrometry, including the expected motion (curved solid line) if it is an unrelated background object; each epoch is labelled by a number 1–6; a dashed line connects the star to each epoch data point; error bars, ±1s.d. Planet e is confirmed as bound to HR 8799, and it is moving at 46 ± 10 mas yr^{-1} anticlockwise. In the main figure, the orbits of the giant planets of our Solar System (Jupiter, Saturn, Uranus and Neptune) are drawn to scale (light grey circles). With a period of ~50 years, the orbit of HR 8799e will be rapidly constrained by future observations; at our current measurement accuracy, it will be possible to measure orbital curvature after only 2 years.

HR 8799e is located very near planets c and d in a K_s versus $K_s - L'$ colour–magnitude diagram, suggesting that all three planets have similar spectral shapes and bolometric luminosities. We, therefore, adopt the same luminosity for these three planets; however, given the larger photometric error-bars and sparse wavelength coverage associated with planet e, we have conservatively assigned to it a luminosity error (Table 1) twice as large as those for planets b, c and d[1]. This luminosity estimate is consistent with empirically calibrated bolometric corrections for brown dwarfs[7], although such corrections may be ill-suited for young planets with distinct spectra and colours. Using the two overlapping age ranges outlined above and the evolutionary models described in the HR 8799bcd discovery article[1], we estimate the mass of planet e to be 7^{+3}_{-2} M_{Jup} (30 Myr) and 10^{+3}_{-3} M_{Jup} (60 Myr), where M_{Jup} is the mass of Jupiter; see Fig. 3. The broadband photometry of planets b, c, and d provide strong evidence for significant atmospheric cloud coverage, while recent spectroscopy of planets b and c show evidence for non-equilibrium CO/CH_4 chemistry[8–10]. Given the limited wavelength coverage of the discovery images for planet e, it is too early to

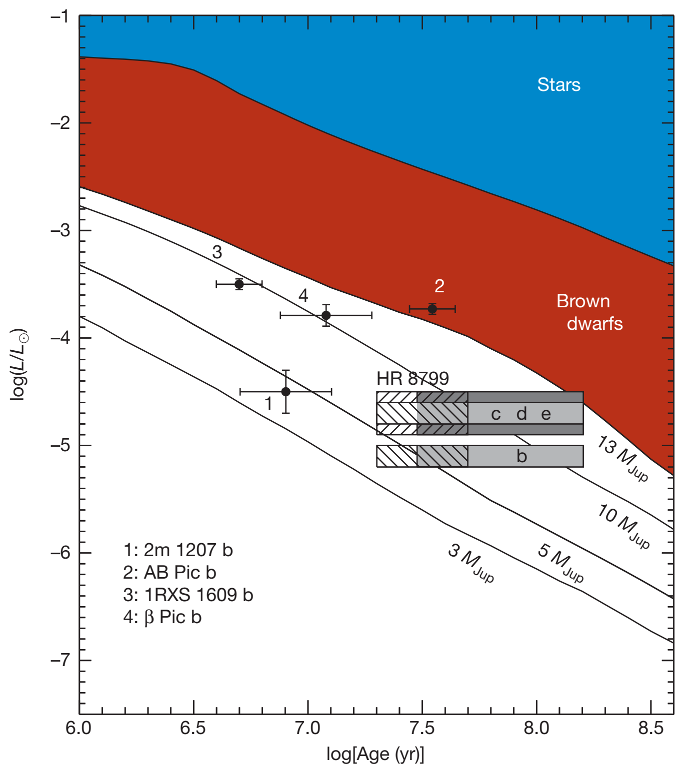


Figure 3 | The mass of HR 8799e from the age–luminosity relationship. Solid lines are luminosity-versus-age tracks for planet evolution models[25] (luminosities are normalized to the solar luminosity, $L_\odot$). Objects above 13 M_{Jup} are typically considered to be outside the planet-mass regime; however, the tail end of the planet distribution found by radial velocity surveys extends above this IAU-defined mass limit[26]. Boxed areas show adopted luminosity ranges (±1s.d.) and estimated age ranges for the four HR 8799 planets: cross-hatched boxes show age range 30^{+20}_{-10} Myr; grey boxes show age range 60^{+100}_{-30} Myr; planets c, d and e have similar luminosities, but the luminosity uncertainty for e is larger and indicated by the darker box/opposite hatch. For comparison, the ages and luminosities of four recently imaged planet-mass companions near other stars are indicated (numbered 1–4; see key on figure) showing 1s.d. error bars for the luminosity and estimated age ranges). An asteroseismology study suggested that the HR 8799 system might be as old as ~1 Gyr (ref. 27), but it is highly unlikely that such an old star would have very massive debris belts[21,28]; such an age would also require planetary masses far too high for long-term stability[13]. The older age also requires an inclination of the stellar pole relative to the line of sight of ~50°, inconsistent with the nearly face-on planetary system and the ~25° inclination upper limit measured from Spitzer images of the outer dust halo[4]. Mass estimates based on any existing evolutionary model at ages as young as 20–30 Myr suffer from unconstrained initial formation conditions; the masses presented here could be underestimated if the planets formed by core-accretion, though 'cold start' core-accretion models[29] do not reproduce the observed luminosity for any combination of mass and age. While this additional uncertainty can lead temporarily to ambiguity about the planets' masses and formation history (core-accretion or gravitation instability), it does highlight the importance of discovering and following in orbit planet-mass companions at ages when formation processes are important.

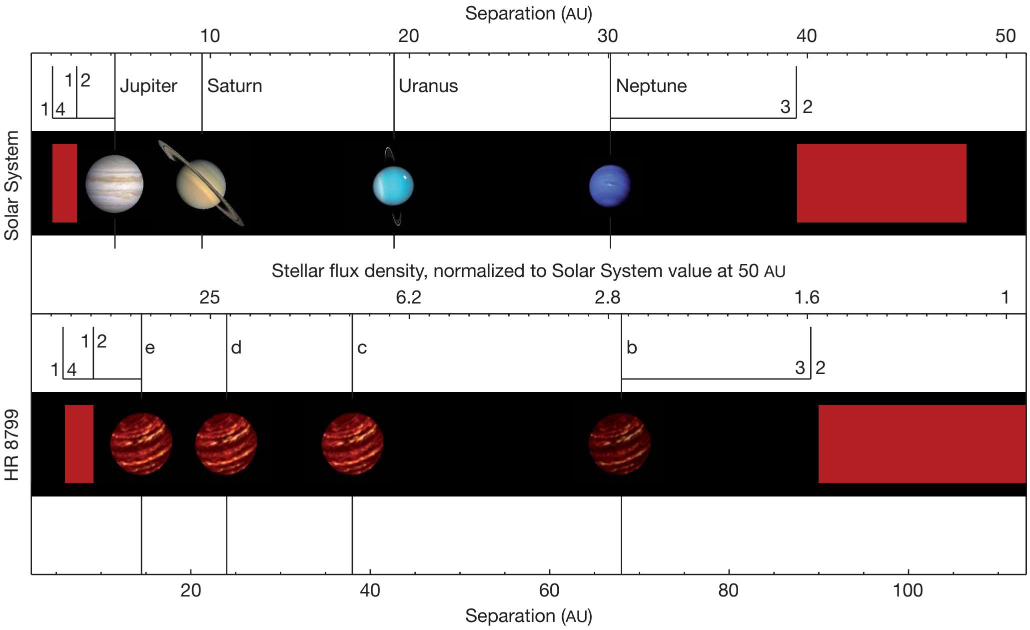


Figure 4 | Comparison of HR 8799 and our Solar System. Top, Solar System; bottom, the HR 8799 system. HR 8799 infrared data indicate the existence of an asteroid belt analogue located at 6–15 AU (we have moved the estimated outer edge of this belt to 10 AU because of planet e's estimated chaotic region[30]), a Edgeworth-Kuiper-belt-like debris disk at >90 AU and a small particle halo extending up to 1,000 AU (ref. 4). The red shaded regions represent the locations of the inner and outer debris belts in both systems (the Solar System Oort comet cloud and the HR 8799 halo are not shown). The horizontal axis of the HR 8799 plot is compressed by the square root of the ratio of the luminosity of HR 8799 ($4.92 \pm 0.41\,L_{\odot}$) to that of the Sun to show both systems over the same equilibrium temperature range. Given the current apparent separations of the four planets of HR 8799 and the preferred locations of the inner warm debris disk and the inner edge of the outer cold disk (90 AU)[4], then (1) the indicated 4:1 and 2:1 period resonances between the inner/outer edge of the warm debris belt and planet e, and (2) a 3:2 mean motion resonance of b with the inner edge of the outer cold disk, are both consistent with the observations. By analogy, the inner and outer edges of the main asteroid belt of our Solar System are, respectively, in 4:1 and 2:1 mean motion resonances with Jupiter. Many members of the Edgeworth–Kuiper belt, including Pluto, are in a 3:2 mean motion orbital resonance with Neptune. Solar System planet images are from NASA; HR 8799 artwork is from Gemini Observatory and L. Cook. Planet diameters are not to scale.

say much about the atmospheric properties of this particular planet; however, given that its near-infrared colour is similar to those of the other three planets, we can anticipate similar cloud structure and chemistry for planet e.

Stability analyses[11–13] have shown that the original three-planet system may be in a mean motion period resonance with an upper limit on planetary masses of $\sim$20 $M_{\rm Jup}$ assuming an age of up to 100 Myr. With the discovery of a fourth planet, we revisit the stability of this system. We searched for stable orbital configurations with the HYBRID/Mercury package[14] using the 30-Myr (5, 7, 7 and 7 $M_{\rm Jup}$ for b, c, d and e respectively) and 60-Myr (7, 10, 10 and 10 $M_{\rm Jup}$) masses. In our preliminary search, we held the parameters for b, c and d fixed to those matching either the single resonance (1:2 resonance between planets c and d only) or double resonance (1:2:4 resonance between planets b, c, and d) stable solutions found to date[13], but allowed the parameters for e to vary within the regime allowed by our observations. On the basis of the single-resonance configuration and using the 30-Myr masses, in 100,000 trials seven solutions for e were found that are stable for at least 160 Myr (the maximum estimated age of the system), and an additional five solutions were found that are stable for over 100 Myr. All maximally stable solutions have a semimajor axis of $\sim$14.5 AU, with planets c, d and e in a 1:2:4 resonance (planet b not in resonance). A set of 100,000 trials was also performed using the 60-Myr masses, but only two solutions were found that are stable for over 100 Myr, each of which requires a semimajor axis of $\sim$12.5 AU, 4σ away from our astrometry. This is suggestive that a younger system age and lower planet masses are preferred, although a much more thorough search of parameter space is required (see the Supplementary Information for tables of stable solutions).

The mechanism for the formation of this system is unclear. It is challenging for gravitational-instability fragmentation to occur at $a < 20$–40 AU (refs 15,16)—ruling that mechanism out for *in situ* formation of planet e. In addition, disk instability mechanisms preferentially form objects more massive than these planets[16,17]. If the HR 8799 system represented low-mass examples of such a population, brown-dwarf companions to young massive stars would be plentiful. Nearby young star surveys[18–20] and our nearly complete survey of 80 stars with similar masses and ages to HR 8799 have discovered no such population of brown dwarf companions. HR 8799e and possibly d are close enough to the primary star to have formed by bottom-up accretion *in situ*[15], but planets b and c are located where the collisional timescale is conventionally thought to be too low for core accretion to form giant planets before the system's gas is depleted. A hybrid process with different planets forming through different mechanisms cannot be ruled out, but seems unlikely with the similar masses and dynamical properties of the four planets. It is possible that one mechanism dominated the other and the planets later migrated to their current positions. The HR 8799 debris disk is especially massive for a star of its age (or for any older main sequence star[21]), which could indicate an extremely dense protoplanetary disk. Such a disk could have induced significant migration, moving planets formed by disk-instability inward, or the disk could have damped the residual eccentricity from multi-planet gravitational interactions that moved core-accretion planets outward. The massive debris disk and the lack of higher-mass analogues to this system do suggest that HR 8799 represents the high-mass end of planet formation.

The HR 8799 system does show interesting similarities with our Solar System; all giant planets are located past the estimated snow line of each system ($\sim$2.7 AU for the Solar System, $\sim$6 AU for HR 8799), and the debris belts of each system are located at similar equilibrium temperatures (Fig. 4). With its four massive planets, massive debris belts and large scale, the HR 8799 planetary system is an amazing example of extreme systems that can form around stars.

Received 5 November; accepted 18 November 2010.
Published online 8 December 2010.

1. Marois, C. *et al.* Direct imaging of multiple planets orbiting the star HR 8799. *Science* **322,** 1348–1352 (2008).
2. Kuiper, G. P. On the origin of the Solar System. *Proc. Natl Acad. Sci. USA* **37,** 1–14 (1951).
3. Mizuno, H. Formation of the giant planets. *Prog. Theor. Phys.* **64,** 544–557 (1980).
4. Su, K. Y. L. *et al.* The debris disk around HR 8799. *Astrophys. J.* **705,** 314–327 (2009).
5. Moór, A. *et al.* Nearby debris disk systems with high fractional luminosity reconsidered. *Astrophys. J.* **644,** 525–542 (2006).
6. Torres, C. A. O., Quast, G. R., Melo, C. H. F. & Sterzik, M. F. in *Handbook of Star Forming Regions* Vol. II, *The Southern Sky* (ed. Reipurth, B.) 757–812 (ASP Monograph Publications, MP 005, 2008).
7. Golimowsky, D. A. *et al.* L' and M' photometry of ultracool dwarfs. *Astrophys. J.* **127,** 3516–3536 (2004).
8. Hinz, P. M. *et al.* Thermal infrared MMTAO observations of the HR 8799 planetary system. *Astrophys. J.* **716,** 417–426 (2010).
9. Janson, M., Bergfors, C., Goto, M., Brandner, W. & Lafrenière, D. Spatially resolved spectroscopy of the exoplanet HR 8799 c. *Astrophys. J.* **710,** L35–L38 (2010).
10. Bowler, B. P., Liu, M. C., Dupuy, T. J. & Cushing, M. C. Near-infrared spectroscopy of the extrasolar planet HR 8799 b. *Astrophys. J.* **723,** 850–868 (2010).
11. Goździewski, K. & Migaszewski, C. Is the HR 8799 extrasolar system destined for planetary scattering? *Mon. Not. R. Astron. Soc.* **397,** L16–L20 (2009).
12. Reidemeister, M. *et al.* A possible architecture of the planetary system HR 8799. *Astron. Astrophys.* **503,** 247–258 (2009).
13. Fabrycky, D. C. & Murray-Clay, R. A. Stability of the directly imaged multiplanet system HR 8799: resonance and masses. *Astrophys. J.* **710,** 1408–1421 (2010).
14. Chambers, J. E. A hybrid symplectic integrator that permits close encounters between massive bodies. *Mon. Not. R. Astron. Soc.* **304,** 793–799 (1999).
15. Dodson-Robinson, S. E., Veras, D., Ford, E. B. & Beichman, C. A. The formation mechanism of gas giants on wide orbits. *Astrophys. J.* **707,** 79–88 (2009).
16. Kratter, K. M., Murray-Clay, R. A. & Youdin, A. N. The runts of the litter: why planets formed through gravitational instability can only be failed binary stars. *Astrophys. J.* **710,** 1375–1386 (2010).
17. Stamatellos, D. & Whitworth, A. P. The properties of brown dwarfs and low-mass hydrogen-burning stars formed by disc fragmentation. *Mon. Not. R. Astron. Soc.* **392,** 413–427 (2009).
18. Lafrenière, D. *et al.* The Gemini Deep Planet Survey. *Astrophys. J.* **670,** 1367–1390 (2007).
19. Nielsen, E. & Close, L. M. A uniform analysis of 118 stars with high-contrast imaging: long-period extrasolar giant planets are rare around Sun-like stars. *Astrophys. J.* **717,** 878–896 (2010).
20. Chauvin, G. *et al.* Deep imaging survey of young, nearby austral stars. VLT/NACO near-infrared Lyot-coronagraphic observations. *Astron. Astrophys.* **509,** A52–A68 (2010).
21. Rhee, J. H., Song, I., Zuckerman, B. & McElwain, M. Characterization of dusty debris disks: the IRAS and Hipparcos catalogs. *Astrophys. J.* **660,** 1556–1571 (2007).
22. Marois, C., Lafrenière, D., Doyon, R., Macintosh, B. & Nadeau, D. Angular differential imaging: a powerful high-contrast imaging technique. *Astrophys. J.* **641,** 556–564 (2006).
23. Lafrenière, D., Marois, C., Doyon, R., Nadeau, D. & Artigau, É. A new algorithm for point-spread function subtraction in high-contrast imaging: a demonstration with angular differential imaging. *Astrophys. J.* **660,** 770–780 (2007).
24. Marois, C., Macintosh, B. & Véran, J.-P. Exoplanet imaging with LOCI processing: photometry and astrometry with the new SOSIE pipeline. *Proc. SPIE* **7736,** 77361J–77361J-12 (2010).
25. Baraffe, I., Chabrier, G., Barman, T. S., Allard, F. & Hauschildt, P. H. Evolutionary models for cool brown dwarfs and extrasolar giant planets. The case of HD 209458. *Astron. Astrophys.* **402,** 701–712 (2003).
26. Eggenberger, A. & Udry, S. Detection and characterization of extrasolar planets through Doppler spectroscopy. *EAS Publ. Ser.* **41,** 27–75 (2010).
27. Moya, A. *et al.* Age determination of the HR8799 planetary system using asteroseismology. *Mon. Not. R. Astron. Soc.* **405,** L81–L85 (2010).
28. Gáspár, A. *et al.* The low level of debris disk activity at the time of the late heavy bombardment: a Spitzer study of Praesepe. *Astrophys. J.* **697,** 1578–1596 (2009).
29. Marley, M. S., Fortney, J., Hubickyj, O., Bodenheimer, P. & Lissauer, J. J. On the luminosity of young Jupiters. *Astrophys. J.* **655,** 541–549 (2007).
30. Malhotra, R. in *Solar System Formation and Evolution* (eds Lazzaro, D., Vieira Martins, R., Ferraz-Mello, S. & Fernandez, J.) 37–63 (ASP Conf. Ser. Vol. 149, 1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Keck staff, particularly H. Lewis, B. Goodrich and J. Lyke, for support with the follow-up observations. We thank G. Laughlin and D.C. Fabrycky for discussions. Portions of this research were performed under the auspices of the US Department of Energy by LLNL and also supported in part by the NSF Center for Adaptive Optics. We acknowledge support by NASA grants to UCLA, LLNL and Lowell Observatory. The data were obtained at the W.M. Keck Observatory. This publication makes use of data products from the Two Micron All Sky Survey and the SIMBAD database.

Author Contributions The authors contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.M. (christian.marois@nrc-cnrc.gc.ca).

doi:10.1038/nature09716

EXTRASOLAR PLANETS

A giant surprise

The discovery of an inner giant planet in the unusually massive solar system around the star HR 8799 creates an ensemble of planets that is difficult to explain with prevailing theories of planet formation.

LAIRD CLOSE

The solar system around the star HR 8799 should not exist. This system is unlike any other known: it is a massive system that has multiple massive planets, with each giant planet containing many times the mass of all the planets in our Solar System combined. However, in a paper published online in *Nature* today, Marois and collaborators[1] present new images of HR 8799 in which yet another equally massive planet is visible.

Previous work[2] had imaged three planets around HR 8799, and now we have the surprise discovery of a fourth, HR 8799e, an inner, massive planet (about 10 Jupiter masses) located some 14.5 astronomical units from the star (1 AU is the average distance from Earth to the Sun). One might question the importance of the discovery of another extrasolar planet when more than 500 are known. But the HR 8799 system is the only solar system known to have multiple outer planets (the other three planets, HR 8799b, HR 8799c and HR 8799d, orbit respectively at approximately 68, 38 and 24 AU from the host star, and have estimated masses of about 7, 10 and 10 Jupiters).

As HR 8799 is the only known example of a wide (greater than 25 AU) solar system with multiple giant planets, astronomers were curious to know whether the star's planets could have formed by gravitational collapse[3] — one of the most popular theories of outer-planet formation. This theory posits that outer giant planets form from the fragmentation of the disk of gas and dust that develops around stars when they are young. In a process rather like the way binary stars form, a gravitational instability in the disk fragments it and quickly (on a timescale of 10,000 years) leads to the formation of gas-giant planets[3]. But the discovery of an inner planet such as HR 8799e at 14.5 AU poses a tricky puzzle. At this distance, the disk was neither cold enough nor rotating slowly enough to fragment and undergo gravitational collapse *in situ* to form HR 8799e[3].

To explain the formation of this latest planet, Marois *et al.*[1] appeal to the dominant theory of giant-planet formation: a slower process than gravitational collapse (about 3.5 million years at a distance of 10 AU) in which solid dust grains conglomerate into solid cores of tens of Earth masses and then gravitationally accrete disk gas to grow to Jupiter masses. Such a 'core-accretion' process itself is only marginally fast enough at 14.5 AU to build up HR 8799e's roughly 10 Jupiter masses before the disk gas accretes onto the star in less than 10 Myr. This formation timescale problem[3] becomes even more vexing if one considers that, at about 2.6 times the distance HR 8799e is from the host star, HR 8799c would require about 20 times longer (more than about 200 Myr) to grow to the same mass at 38 AU — long after the disk has lost all its gas. What's more, at 68 AU, HR 8799b's formation is truly problematic, requiring an even longer timescale (many times the age of the star) to have formed *in situ* by core accretion. Hence, neither of the two favoured theories of giant-planet formation can explain how all the planets around HR 8799 formed: HR 8799e is too close to have formed by gravitational collapse, and HR 8799c and HR 8799b are too far out to have formed by core accretion (Fig. 1).

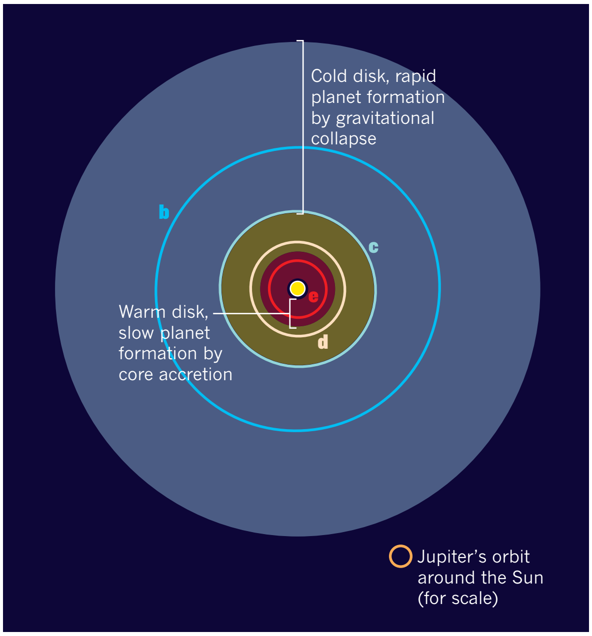


Figure 1 | The HR 8799 planetary system. When star HR 8799 formed, a massive circumstellar disk of gas and dust probably existed from which the star's four massive planets formed; the planets' approximate current orbits are overlaid and labelled b–e. The outer part of the disk was very cold and rotated slowly, and so might have collapsed through gravitational instabilities to quickly form outer planets such as 'b'. The newly discovered 'e' planet[1] is in a very different zone, where the disk was much warmer and the planet is likely to have formed in a slow, two-step 'core-accretion' process. Neither theory of planet formation — gravitational collapse or core accretion — can explain the whole family of four planets.

Perhaps all of these massive planets formed at much larger distances (more than at least 50 AU) by the gravitational collapse of an unusually massive disk and then migrated quickly inwards to their current positions, somehow sweeping into a dynamically stable set of 1:2:4 orbital resonances[1] (where, for every one orbit of planet c, there are two of d and four of e). This does not really help the situation, however, because it is unlikely that such a massive planet as HR 8799e could have migrated from about 50 to 14.5 AU by means of tidal torques from the residual gas that had not been used to build up the planets. The converse theory, by which the planets all form through core accretion within about 10 AU and then slowly move outwards by scattering lesser objects (planetesimals) inwards, is also problematic because there is probably too limited a reservoir of planetesimals to move a 7-Jupiter-mass object such as HR 8799b outwards some 58 AU. So, despite having a clear view of the system — thanks to the power of adaptive-optics systems and large ground-based telescopes — we cannot currently explain how all four planets formed in a coherent, coeval fashion.

A key strength of direct imaging is that photons can be collected from these self-luminous young planets as they contract, allowing the planetary spectra to be observed (to calculate temperatures and luminosities). The observed brightness of HR 8799b in direct images is much lower than would be expected from its observed temperature, given that evolutionary models indicate that HR 8799b must have a radius larger than that of Jupiter[1,2,4].

doi:10.1038/nature09716

This 'under-luminosity' problem is typical of around half of the extrasolar planets imaged to date. One possible explanation is that dusty, thick, planetary-scale high-latitude cloud 'bands' absorb/scatter light when viewing a young planet over its pole. For example, the 'under-luminous' planets in the HR 8799 system are probably being viewed close to 'pole-on'[5], perhaps leading to less light emitted in the direction of Earth. By contrast, 'edge-on' giant planets, such as β-Pictoris b[6], look brighter because light streams freely from the brighter equatorial regions between the dark cloud bands. Clearly, further theoretical (and direct imaging) work will be needed to identify the ultimate cause of this under-luminosity problem.

The future holds much promise for more surprises in the field of direct imaging of extrasolar planets. However, it seems unlikely that any other massive outer planets will be found around HR 8799[7]. There is always a chance, though, that low-mass terrestrial planets lie within the star's 10-AU-radius 'asteroid' belt. The next chapter in this story will soon be written by even more powerful ground-based, adaptive-optics imagers[8,9] and, let us hope, by more powerful pathfinding, space-based planet- and disk-imaging telescopes[10]. These pathfinders should eventually lead to a terrestrial-planet-finding telescope even capable of taking spectra of Earth-like planets. Such an achievement could address one of the most pivotal questions in science: how common are truly Earth-like planets and life in our Universe? ■

Laird Close *is in the Department of Astronomy and Steward Observatory, University of Arizona, Tucson, Arizona 85721, USA.*
e-mail lclose@as.arizona.edu

1. Marois, C., Zuckerman, B., Konopacky, Q. M., Macintosh, B. & Barman, T. *Nature* doi:10.1038/nature09684 (2010).
2. Marois, C. *et al. Science* **322,** 1348–1352 (2008).
3. Nero, D. & Bjorkman, J. E. *Astrophys. J.* **702,** L163–L167 (2009).
4. Bowler, B. P., Liu, M. C., Dupuy, T. J. & Cushing, M. C. *Astrophys. J.* **723,** 850–868 (2010).
5. Moro-Martín, A., Rieke, G. H. & Su, K. Y. L. *Astrophys. J.* **721,** L199–L202 (2010).
6. Lagrange, A.-M. *et al. Science* **329,** 57–59 (2010).
7. Close, L. M. & Males, J. R. *Astrophys. J.* **709,** 342–348 (2010).
8. Beuzit, J.-L. *et al. Astron. Soc. Pacif. Conf. Ser.* **430,** 231 (2010).
9. Macintosh, B. A. *et al. Proc. SPIE* **7015,** 701518 (2008).
10. Green, T. P., Schneider, G. & EXCEDE Mission Team *Bull. Am. Astron. Soc.* **39,** 975 (2007).